

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:38:48 ; Search time 9.06061 Seconds
(without alignments)
1368.719 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDSNT.....GRLSQLHTFEKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1647	100.0	299	1 MTL5_HUMAN	Q9Y415 homo sapien
2	1251	76.0	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
3	109.5	6.6	769	1 ITB2_HUMAN	P05107 homo sapien
4	106.5	6.5	3718	1 LMA5_MOUSE	Q61001 mus musculus
5	103.5	6.3	1786	1 LMB1_MOUSE	P02469 mus musculus
6	103	6.3	769	1 ITB2_PIG	P53714 sus scrofa
7	102	6.2	1790	1 LMB1_DROME	P11046 drosophila
8	101.5	6.2	396	1 TFH2_MOUSE	Q9j1b4 mus musculus
9	101	6.1	1700	1 BAR3_CHITE	Q03376 chironomus
10	101	6.1	3011	1 POLG_HCV1	P26664 h genome po
11	100.5	6.1	760	1 E2_DROME	P42124 drosophila
12	100	6.1	3595	1 LMA5_HUMAN	O15230 homo sapien
13	99	6.0	787	1 ITB6_MOUSE	Q920t9 mus musculus
14	98.5	6.0	3106	1 LMA2_MOUSE	Q60675 mus musculus
15	98.5	6.0	5376	1 ZAN_MOUSE	O88799 mus musculus
16	97.5	5.9	328	1 C170_GIALA	P15799 giardia lam
17	97.5	5.9	643	1 CD93_RAT	Q9et61 rattus norv
18	97.5	5.9	969	1 PAC4_HUMAN	P29122 homo sapien
19	97.5	5.9	971	1 RECK_HUMAN	O95980 homo sapien
20	97	5.9	395	1 TFH2_HUMAN	Q13888 homo sapien
21	97	5.9	798	1 ITB7_HUMAN	P26010 homo sapien
22	97	5.9	3010	1 POLG_HCVJT	Q00269 h genome po
23	96.5	5.9	290	1 Y070_NPVAC	P41470 autographa
24	96.5	5.9	1576	1 YLK3_CABEL	P41951 caenorhabdi
25	96.5	5.9	3726	1 TRX_DROME	P20659 drosophila
26	96	5.8	798	1 ITB1_XENLA	P12606 xenopus lae
27	96	5.8	1808	1 TENA_CHICK	P10039 gallus gall
28	96	5.8	3010	1 POLG_HCVBK	P26663 h genome po
29	95.5	5.8	747	1 E2H1_HUMAN	Q92800 homo sapien
30	95.5	5.8	1801	1 LMB2_RAT	P15800 rattus norv
31	94	5.7	3011	1 POLG_HCVH	P27958 h genome po
32	93	5.6	3110	1 LMA2_HUMAN	P24043 homo sapien
33	93	5.6	3224	1 RBP2_HUMAN	P49792 homo sapien

34 92.5 5.6 332 1 XYNA_CRYAL
35 92.5 5.6 611 1 LEM2_CANFA
36 92.5 5.6 773 1 MSU2_DROME
37 92.5 5.6 1172 1 AHM2_ARATH
38 92 5.6 798 1 ITB0_XENLA
39 92 5.6 3075 1 LMA1_HUMAN
40 91.5 5.6 4544 1 LRP1_HUMAN
41 90.5 5.5 354 1 VEGD_HUMAN
42 90.5 5.5 747 1 E2H1_MOUSE
43 90.5 5.5 1786 1 LMB1_HUMAN
44 90 5.5 3010 1 POLG_HCVTW
45 89.5 5.4 771 1 ITB2_MOUSE

ALIGNMENTS

RESULT 1

ID MTL5_HUMAN STANDARD; PRT: 299 AA.
AC Q9Y415;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RT early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U86074; AAD24668.1;
CC Genew; HGNC:7446; MTL5.
DR MIM: 604374; -;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH
SQ SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;
Query Match 100.0%; Score 1647; DB 1; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.5e-134;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVICQLKGGTQMLCIDSNTRELKALHVPQYQDQNNYQLQSDVPKPMTALVGRFLPASTK 60
DB 1 MVICQLKGGTQMLCIDSNTRELKALHVPQYQDQNNYQLQSDVPKPMTALVGRFLPASTK 60
QY 61 LNLITQLEGALPSVVVNGSAFPGSGTLPKPKITLAGYCDCAFSGDFCNCCNCCNLL 120
DB 61 LNLITQLEGALPSVVVNGSAFPGSGTLPKPKITLAGYCDCAFSGDFCNCCNCCNLL 120
QY 121 HHDIERFRAIKACLRNPFAFPKIGKGLGNVYKPKHNGKNCRRSGCKLKNYCEYEAQI 180
DB 121 HHDIERFRAIKACLRNPFAFPKIGKGLGNVYKPKHNGKNCRRSGCKLKNYCEYEAQI 180

```

Db 121 HHDIERFAKRAICLRNPEAFQPKIGKQGLGNVKNQHNKGNCRSRGCKLKNYCECYEQAQI 180
Qy 181 MCSSTICKIGCKNYEESPERKTLASMPNMYMTGGLEGSHYLPPTKFSGLPFRSHDRPSS 240
Db 181 MCSSTICKIGCKNYEESPERKTLASMPNMYMTGGLEGSHYLPPTKFSGLPFRSHDRPSS 240
Qy 241 CISWVVEATCACLAAQEEAEKHCSCKLAEDMLEEFGRCLSOILHTEFKSKGLKME 299
Db 241 CISWVVEATCACLAAQEEAEKHCSCKLAEDMLEEFGRCLSOILHTEFKSKGLKME 299

RESULT 2
MTL5_MOUSE
ID ITB2_MOUSE STANDARD; PRT; 295 AA.
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE tesmin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).
GN MTL5
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77383; AAD24667.1; -
DR MGD; MGI:1340029; MTL5.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Spermatogenesis.
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Query Match 76.0%; Score 1251; DB 1; Length 295;
Best Local Similarity 76.3%; Pred. No. 2.5e-100;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICOLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYLOSDVPKPMFTALYGRFLPASTK 60
Db 1 MVICOLKGAQMLCIDNGARELALHLLPQYDQSSFPQSELPKMTTLVGRLLPVPK 60
Qy 61 LNLITQLEGALPSVVGSAFSPGSTLPKPKITLAGYCDGFCFAGDPCNCCNCCNL 120
Db 61 LNLITQVNGALPSAVNGAFAFSPALQGPCKITLSGYCDGFCFSGDFGSCS----CNL 116
Qy 121 HHDIERFAKRAICLRNPEAFQPKIGKQGLGNVKNQHNKGNCRSRGCKLKNYCECYEQAQI 180
Db 117 RHELFKFAKRAICLRNPEAFQPKIGKQGLGNVKNQHNKGNCRSRGCKLKNYCECYEQAQI 176
Qy 181 MCSSTICKIGCKNYEESPERKTLASMPNMYMTGGLEGSHYLPPTKFSGLPFRSHDRPSS 240

```

```

Db 177 MCSSTICKIACKNYEESPERKMLASTHYMEPGDFESSHLYSPAKFSPPKLRNQAQS 236
Qy 241 CISWVVEATCACLAAQEEAEKHCSCKLAEDMLEEFGRCLSOILHTEFKSKGLKME 299
Db 237 CISWVVEATCACLAAQEEAEKHCSPSLAEQMLEEFGRCLSOILHIEFKSKGLKIE 295

RESULT 3
ITB2_HUMAN
ID ITB2_HUMAN STANDARD; PRT; 769 AA.
AC P05107; Q16418;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CR3/p150,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit).
GN ITGB2 OR CD18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87131080; PubMed=3028646;
RA Kishimoto T.K., O'Connor K., Lee A., Roberts T.M., Springer T.A.;
RT "Cloning of the beta subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family.";
RL Cell 48:681-690(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92077153; PubMed=1683838;
RA Weitzman J.B., Wells C.E., Wright A.H., Clark P.A., Law S.K.A.;
RT "The gene organisation of the human beta 2 integrin subunit (CD18).";
RL FEBS Lett. 294:97-103(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.B., Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P., Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 9-769 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Spleen;
RX MEDLINE=87246525; PubMed=2954816;
RA Law S.K.A., Gagnon J., Hildreth J.E., Wells C.E., Willis A.C., Wong A.J.;
RT "The primary structure of the beta-subunit of the cell surface adhesion glycoproteins LFA-1, CR3 and p150,95 and its relationship to the fibronectin receptor.";
RL EMBO J. 6:915-919(1987).
RN [5]
RP SEQUENCE OF 347-355 FROM N.A., AND VARIANTS LAD1 SER-351 AND TRP-586.
RX MEDLINE=92147694; PubMed=1346613;
RA Nelson C., Rabb H., Arnaout M.A.;
RT "Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing and a missense mutation in a conserved region of CD18 impair cell surface expression of beta 2 integrins.";
RL J. Biol. Chem. 267:3351-3357(1992).
RN [6]

```


RP VARIANTS LAD1 PRO-149 AND ARG-169.
 RX MEDLINE-90293696; PubMed-1694220;
 RA Wardlaw A.J., Hibbs M.L., Stacker S.A., Springer T.A.;
 RT "Distinct mutations in two patients with leukocyte adhesion
 deficiency and their functional correlates";
 RL J. Exp. Med. 172:335-345(1990).
 RN [17]
 RP VARIANT LAD1 ARG-169.
 RX MEDLINE-92324303; PubMed-1352501;
 RA Corbi A., Vara A., Ursa A., Rodriguez M.C.G., Fontan G.,
 RA Sanchez-Madrid F.;
 RT "Molecular basis for a severe case of leukocyte adhesion deficiency.";
 RL Eur. J. Immunol. 22:1877-1881(1992).
 RN [18]
 RP VARIANTS LAD1 THR-196 AND CYS-593.
 RX MEDLINE-90187099; PubMed-1968911;
 RA Arnout M.A., Dana N., Gupta S.K., Tenen D.G., Fathallah D.M.;
 RT "Point mutations impairing cell surface expression of the common beta
 subunit (CD18) in a patient with leukocyte adhesion molecule
 (Leu-CAM) deficiency.";
 RL J. Clin. Invest. 85:977-981(1990).
 RN [19]
 RP VARIANT LAD1 LEU-178.
 RX MEDLINE-92184805; PubMed-1347532;
 RA Back L.L., Kwok W.W., Hickstein D.D.;
 RT "Identification of two molecular defects in a child with leukocyte
 adherence deficiency.";
 RL J. Biol. Chem. 267:5482-5487(1992).
 RN [10]
 RP VARIANT LAD1 ASN-128.
 RX MEDLINE-92272746; PubMed-1590804;
 RA Matsuura S., Kishi F., Tsukahara M., Nunci H., Matsuda I.,
 RA Kobayashi K., Kajii T.;
 RT "Leukocyte adhesion deficiency: identification of novel mutations in
 two Japanese patients with a severe form.";
 RL Biochem. Biophys. Res. Commun. 184:1460-1467(1992).
 RN [11]
 RP VARIANT LAD1 SER-284.
 RX MEDLINE-93312347; PubMed-7686755;
 RA Back L.A., Kerkering M., Baker D., Bauer T.R., Embree L.J.,
 RA Hickstein D.D.;
 RT "A point mutation associated with leukocyte adhesion deficiency type
 1 of moderate severity";
 RL Biochem. Biophys. Res. Commun. 193:912-918(1993).
 RN [12]
 RP VARIANTS LAD1 PRO-138 AND ARG-273.
 RX MEDLINE-99102253; PubMed-9884339;
 RA Hogg N., Stewart M.P., Scarth S.L., Newton R., Shaw J.M., Law S.K.A.,
 RA Klein N.;
 RT "A novel leukocyte adhesion deficiency caused by expressed but
 nonfunctional beta2 integrins Mac-1 and LFA-1.";
 RL J. Clin. Invest. 103:97-106(1999).
 CC -1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
 CC ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE
 CC RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT
 CC AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE
 CC SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2
 CC REGONIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN
 CC ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-
 CC D/BETA 2 IS A RECEPTOR FOR ICAM3 AND VCAM1.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2
 CC ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DISEASE: DEFECTS IN ITGB2 ARE THE CAUSE OF LEUKOCYTE ADHESION
 CC DEFICIENCY, TYPE I (LAD1). LAD1 PATIENTS HAVE RECURRENT BACTERIAL
 CC INFECTIONS AND THEIR LEUKOCYTES ARE DEFICIENT IN A WIDE RANGE OF
 CC ADHESION-DEPENDENT FUNCTIONS.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD18 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd18.htm".
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; Y00057; CAA68266.1;
 DR EMBL; X64072; CAA45427.1;
 DR EMBL; X64073; CAA45427.1; JOINED.
 DR EMBL; X64074; CAA45427.1; JOINED.
 DR EMBL; X64075; CAA45427.1; JOINED.
 DR EMBL; X64076; CAA45427.1; JOINED.
 DR EMBL; X64077; CAA45427.1; JOINED.
 DR EMBL; X64078; CAA45427.1; JOINED.
 DR EMBL; X64079; CAA45427.1; JOINED.
 DR EMBL; X64080; CAA45427.1; JOINED.
 DR EMBL; X64081; CAA45427.1; JOINED.
 DR EMBL; X64082; CAA45427.1; JOINED.
 DR EMBL; X64083; CAA45427.1; JOINED.
 DR EMBL; X63924; CAA45427.1; JOINED.
 DR EMBL; X63925; CAA45427.1; JOINED.
 DR EMBL; X63926; CAA45427.1; JOINED.
 DR EMBL; AL163300; CAB90553.1;
 DR EMBL; M15395; AAA59490.1;
 DR EMBL; S81234; AAB21404.1;
 DR PIR; A25967; IJHULM.
 DR PIR; A29265; A29265.
 DR HSSP; P05106; IJUV2.
 DR Genew; HGNC:6155; ITGB2.
 DR MIM; 600065;
 DR MIM; 116920;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002369; Integrin_B.
 DR InterPro; IPR001169; Integrin_beta_C.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00362; Integrin_B; 1.
 DR PRINTS; PR01186; INTEGRINB.
 DR Prodom; PD001811; Integrin_B; 1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA; 3.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
 KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KW Repeat; Signal; Disease mutation.
 FT SIGNAL 1 22
 FT CHAIN 23 769 INTEGRIN BETA-2.
 FT DOMAIN 23 700 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 701 723 POTENTIAL.
 FT DOMAIN 724 769 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 134 363 VWFA-LIKE.
 FT DOMAIN 449 617 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 449 496 I.
 FT REPEAT 497 540 II.
 FT REPEAT 541 581 III.
 FT REPEAT 582 617 IV.
 FT DISULFID 25 447 BY SIMILARITY.
 FT DISULFID 33 43 BY SIMILARITY.
 FT DISULFID 36 73 BY SIMILARITY.
 FT DISULFID 46 62 BY SIMILARITY.
 FT DISULFID 191 198 BY SIMILARITY.

Query Match Score 6.6%; DB 1; Length 769;
 Best Local Similarity 22.3%; Pred. No. 0.069;
 Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MWICQLKGGTQMLCIDNSTRELKALHLVPOYQDQNNYLSQDVPKPMVTLVGRFLPASTK 60
 DB 436 IVTVQVLPOQECRCRQSDRSL--CH-----GKGFLECGICRCDTGYIGNKCECQTQ 486

QY 61 LNLITQOLEGA-----LPSVVGSAFFSGSTLPQPKITLAGYCDCA----- 103
 Db 487 -GRSSQLEGGSCRDNNISICGLGDCVGCGLCHTSDVPG--KLIQYQYCECDTINCR 543
 QY 104 -SGDFCINN-----CNCNCCNLLHDIERF-----KAIKACLRNPEAFQPKIGKQGLGN 152
 Db 544 YNGQVCGPGRGLGFCGKRC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR--- 595
 QY 153 VKPQHNGKNCRRSGCLKNYCEVQAQIM-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCEHSGQLPLCQECPCGSPCGKYISACELKFEKGP 638

RESULT 4
 LMA5_MOUSE
 ID LMA5_MOUSE STANDARD; PRT; 3718 AA.
 AC Q61001; Q9JHQ6;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-5 chain precursor.
 GN LMA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] SEQUENCE OF 1-92 FROM N.A.
 RA Timpl R., Sakaki T.;
 RT "Completion of the N-terminal sequence of the murine Laminin alpha 5 chain.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2] SEQUENCE OF 84-3718 FROM N.A.
 RC STRAIN-C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=96081906; PubMed=7499364;
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread expression in adult mouse tissues.";
 RL J. Biol. Chem. 270:28523-28526(1995).
 RN [3] REVISTONS.
 RP Miner J.H., Lewis R.M., Sanes J.R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, Laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.
 CC -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).
 CC -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT AND SKIN.
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ293593; CAB99255.1; -
 DR EMBL; U37501; AAC53430.1; -
 DR HSSP; P02468; ITLE.
 DR MGI; MGI:105382; Lama5.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; Laminin.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00052; laminin_B; 1.
 DR Pfam; PF00053; laminin_EGF; 19.
 DR Pfam; PF00054; laminin_G; 2.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR ProDom; PD002082; Lamn1; 1.
 DR ProDom; PD003031; Laminin_B; 1.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 40 POTENTIAL.
 FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
 FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 305 363 LAMININ EGF-LIKE 1.
 FT DOMAIN 364 433 LAMININ EGF-LIKE 2.
 FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
 FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
 FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
 FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
 FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
 FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
 FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
 FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
 FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
 FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
 FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
 FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
 FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
 FT DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
 FT DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
 FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
 FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
 FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
 FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
 FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
 FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
 FT DOMAIN 2169 2735 DOMAIN II AND I.
 FT DOMAIN 2736 LAMININ G-LIKE 1.
 FT DOMAIN 2947 3119 LAMININ G-LIKE 2.
 FT DOMAIN 3128 3296 LAMININ G-LIKE 3.
 FT DOMAIN 3337 3511 LAMININ G-LIKE 4.
 FT DOMAIN 3518 3689 LAMININ G-LIKE 5.
 FT DOMAIN 2205 COILED COIL (POTENTIAL).
 FT DOMAIN 2330 COILED COIL (POTENTIAL).
 FT DOMAIN 2604 COILED COIL (POTENTIAL).
 FT DOMAIN 2639 COILED COIL (POTENTIAL).
 FT SITE 1723 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1839 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 305 BY SIMILARITY.
 FT DISULFID 307 BY SIMILARITY.
 FT DISULFID 329 BY SIMILARITY.
 FT DISULFID 341 BY SIMILARITY.
 FT DISULFID 364 BY SIMILARITY.
 FT DISULFID 366 BY SIMILARITY.
 FT DISULFID 401 BY SIMILARITY.
 FT DISULFID 431 BY SIMILARITY.
 FT DISULFID 434 BY SIMILARITY.
 FT DISULFID 445

CC CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
 CC CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 DR EMBL; M15525; AAA39407.1; ALT_INIT.
 DR EMBL; X05212; CAA28839.1; -;
 DR PIR; A26413; MMMSBI.
 DR HSP; P02468; 1KLO.
 DR MGD; MGI:96743; Lamb1-1.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00053; laminin_EGF; 13.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR ProDom; PD002082; LamNT; 1.
 DR SMART; SM00180; EGF_Lam; 11.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 9.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.
 FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 271 334 LAMININ EGF-LIKE 1.
 FT DOMAIN 335 397 LAMININ EGF-LIKE 2.
 FT DOMAIN 398 457 LAMININ EGF-LIKE 3.
 FT DOMAIN 458 509 LAMININ EGF-LIKE 4.
 FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 541 772 LAMININ DOMAIN IV.
 FT DOMAIN 773 820 LAMININ EGF-LIKE 6.
 FT DOMAIN 821 866 LAMININ EGF-LIKE 7.
 FT DOMAIN 867 916 LAMININ EGF-LIKE 8.
 FT DOMAIN 917 975 LAMININ EGF-LIKE 9.
 FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.
 FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.
 FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.
 FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.
 FT DOMAIN 1179 1397 DOMAIN II.
 FT DOMAIN 1398 1430 DOMAIN ALPHA.
 FT DOMAIN 1431 1786 DOMAIN I.
 FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).
 FT DOMAIN 1368 1388 COILED COIL (POTENTIAL).
 FT DOMAIN 1448 1778 COILED COIL (POTENTIAL).
 FT DISULFID 271 280 BY SIMILARITY.
 FT DISULFID 273 298 BY SIMILARITY.
 FT DISULFID 300 309 BY SIMILARITY.
 FT DISULFID 312 332 BY SIMILARITY.
 FT DISULFID 335 344 BY SIMILARITY.
 FT DISULFID 337 362 BY SIMILARITY.
 FT DISULFID 365 374 BY SIMILARITY.
 FT DISULFID 377 395 BY SIMILARITY.
 FT DISULFID 398 411 BY SIMILARITY.
 FT DISULFID 400 426 BY SIMILARITY.
 FT DISULFID 428 437 BY SIMILARITY.
 FT DISULFID 440 455 BY SIMILARITY.
 FT DISULFID 458 472 BY SIMILARITY.
 FT DISULFID 460 479 BY SIMILARITY.
 FT DISULFID 481 490 BY SIMILARITY.
 FT DISULFID 493 507 BY SIMILARITY.
 FT DISULFID 773 785 BY SIMILARITY.
 FT DISULFID 775 792 BY SIMILARITY.
 FT DISULFID 794 803 BY SIMILARITY.
 FT DISULFID 806 818 BY SIMILARITY.

FT DISULFID 821 833 BY SIMILARITY.
 FT DISULFID 823 840 BY SIMILARITY.
 FT DISULFID 851 864 BY SIMILARITY.
 FT DISULFID 854 866 BY SIMILARITY.
 FT DISULFID 867 883 BY SIMILARITY.
 FT DISULFID 869 895 BY SIMILARITY.
 FT DISULFID 886 895 BY SIMILARITY.
 FT DISULFID 898 914 BY SIMILARITY.
 FT DISULFID 917 933 BY SIMILARITY.
 FT DISULFID 919 944 BY SIMILARITY.
 FT DISULFID 946 953 BY SIMILARITY.
 FT DISULFID 958 973 BY SIMILARITY.
 FT DISULFID 976 990 BY SIMILARITY.
 FT DISULFID 978 997 BY SIMILARITY.
 FT DISULFID 1000 1009 BY SIMILARITY.
 FT DISULFID 1012 1025 BY SIMILARITY.
 FT DISULFID 1084 1096 BY SIMILARITY.
 FT DISULFID 1086 1103 BY SIMILARITY.
 FT DISULFID 1105 1114 BY SIMILARITY.
 FT DISULFID 1117 1129 BY SIMILARITY.
 FT DISULFID 1132 1144 BY SIMILARITY.
 FT DISULFID 1134 1151 BY SIMILARITY.
 FT DISULFID 1153 1162 BY SIMILARITY.
 FT DISULFID 1165 1176 BY SIMILARITY.
 FT DISULFID 1179 1179 INTERCHAIN (PROBABLE).
 FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).
 FT DISULFID 1185 1785 INTERCHAIN (PROBABLE).
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1041 1041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1195 1195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1279 1279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1343 1343 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1533 1533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1542 1542 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1643 1643 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1531 1534 SGNA -> MEMP (IN REF. 2).
 FT CONFLICT 1749 1749 D -> N (IN REF. 2).
 SQ SEQUENCE 1786 AA; 196904 MW; 846671B7BF41A474 CRC64;
 Query Match 6.3%; Score 103.5; DB 1; Length 1786;
 Best Local Similarity 21.4%; Pred. No. 0.55;
 Matches 54; Conservative 29; Mismatches 74; Indels 95; Gaps 15;
 QY 90 PKITLAGYCDCFASGDFCNCC-----NCNCCNNLHHDIERFKAKICIGR 136
 Db 937 PVTQLACVDPGYIGSRCDDCASGFFGNPDSFGGSCPC--QCHHNIDT-----T 985
 QY 137 NPEAFQPKIGK-----GQLGNVKQHNKGCNCRSGCLKNY-----C 173
 Db 986 DPEACDKDTGTCLKLYHTEGDHQLCOYGYGDALQDCKVCYNLTGTVKEHCNSDC 1045
 QY 174 ECEAQMCSICKICIGCKNVEESPERKTLMSNPNYQW---TG---GLEGSHYLPPT- 224
 Db 1046 HCDKATQCCSCLPNVIG-----QNCDR---CAPNTWQLASGTGCGPCNCAHISFGPSC 1096
 QY 225 -KFSG---LPRFSDRRPSSC---ISW-----EVVEATCACLLAQ 257
 Db 1097 NEFTGQCQMPGFG--GRTCEBCEQLFWGDPDVECRACDPRGTETPCDQSTQCVCVE 1155
 QY 258 GEEAEKEHCKSC 269
 Db 1156 GVEGPR--CDKC 1165
 RESULT 6
 ITB2_PIG
 ID ITB2_PIG STANDARD; PRT; 769 AA.
 AC PS3714;

DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CD3/PI50,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit).
 DE subunit).
 GN ITGB2 OR CD18.
 OS Sus scrofa (Plg).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lee J.K., Schock L.B., Rutherford M.S.;
 RT "Molecular cloning and characterization of the porcine CD18 leukocyte adhesion molecule."
 RL xenotransplantation 3:222-230(1996).
 CC -!- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2, ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2 RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND VCAM1.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2 ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U13941; AAB16868.1; .
 DR HSP; P05106; IJV2.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002369; Integrin_B.
 DR InterPro; IPR001169; Integrin_beta_C.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR002035; VWFA.
 DR Pfam; PF00362; Integrin_B; 1.
 DR PRINTS; P00186; INTEGRINB.
 DR ProDom; PD00181; Integrin_B; 1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA; 3.
 DR PROSITE; PS00222; EGF_1; UNKNOWN_2.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
 KW Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 769
 FT DOMAIN 23 700
 FT TRANSMEM 701 723
 FT DOMAIN 724 769
 FT DOMAIN 124 363
 FT DOMAIN 449 617
 FT REPEAT 449 496
 FT REPEAT 497 540
 FT REPEAT 541 581
 FT REPEAT 582 617
 FT DISULFID 25 447
 FT DISULFID 33 43
 FT DISULFID 36 73
 FT DISULFID 46 62
 FT DISULFID 191 198
 FT DISULFID 246 286

FT DISULFID 386 400 BY SIMILARITY.
 FT DISULFID 420 562 BY SIMILARITY.
 FT DISULFID 445 449 BY SIMILARITY.
 FT DISULFID 459 470 BY SIMILARITY.
 FT DISULFID 467 506 BY SIMILARITY.
 FT DISULFID 472 481 BY SIMILARITY.
 FT DISULFID 483 497 BY SIMILARITY.
 FT DISULFID 512 517 BY SIMILARITY.
 FT DISULFID 514 549 BY SIMILARITY.
 FT DISULFID 519 534 BY SIMILARITY.
 FT DISULFID 536 541 BY SIMILARITY.
 FT DISULFID 557 562 BY SIMILARITY.
 FT DISULFID 559 590 BY SIMILARITY.
 FT DISULFID 564 573 BY SIMILARITY.
 FT DISULFID 575 582 BY SIMILARITY.
 FT DISULFID 596 601 BY SIMILARITY.
 FT DISULFID 598 643 BY SIMILARITY.
 FT DISULFID 603 612 BY SIMILARITY.
 FT DISULFID 615 618 BY SIMILARITY.
 FT DISULFID 622 631 BY SIMILARITY.
 FT DISULFID 628 695 BY SIMILARITY.
 FT DISULFID 647 670 BY SIMILARITY.
 FT SITE 397 399 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 50 50 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 642 642 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 769 AA; 84789 MW; FDD606CEEE850449 CRC64;
 Query Match 6.3%; Score 103; DB 1; Length 769;
 Best Local Similarity 21.5%; Pred. No. 0.25;
 Matches 59; Conservative 28; Mismatches 88; Indels 100; Gaps 16;
 QY 14 CIDNSRTRELKALHLVPOYQDNNYLQSDVPKPMALVGRFLPASTKLNLIITQLEGALP 73
 Db 449 CGDSKERTLCG-----NKGSMGCVCRDAGYIGKHCECOTQ-GRSQLEGSCR 498
 QY 74 SVVNGSAFSGSTLPGPKITLAGYCDCFAS-----GDFCNCCNCCNCCNLLHHDE 125
 Db 499 K-DNSSIICSGL-----GDCICGQCVCVCHTSVDPNKKIYGQFC-ECDNMNC-----E 542
 QY 126 RFKAIKACLG-----RNPEAFQPKIGKGQGLGNVKGQHNKGCNRRS--GCLK-- 170
 Db 543 REDG-QVCGEGKRGKLCFCSTCRQSGFE-----GSACQCLKSTQGLNLQ 586
 QY 171 -----NYCEC--YE-----AQIMCSSICKICGCKNYEESPERKTLMSM 206
 Db 587 GVECSGRGRRCNVCQCDFGYQPLCTDCPCQVPCARYAKCAECLKFDTPGFAKNCSAE 646
 QY 207 PNYMTGLEGSHYLPPTKFGSLPRFSDRRPSSC 241
 Db 647 C-----GTTKLPMSMSG--RKCNERDSEGC 670
 RESULT 7
 LMB1_DROME
 ID LMB1_DROME STANDARD; PRT; 1790 AA.
 AC P11046; Q26328; Q9XZT4; Q9VLM6;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin beta-1 chain precursor (Laminin B1 chain).
 GN LAMB1 OR LAMB1 OR CG7123.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=94000382; PubMed=8397815;

RA Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;
 RT "Analysis of the Drosophila gene for the laminin B1 chain";
 RL DNA Cell Biol. 12:573-587(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88210471; PubMed=3365769;
 RC STRAIN=Berkely;
 RD MEDLINE=20196006; PubMed=107311132;
 RE Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.F., Benos P.V., Bertram J., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman K.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RL "The genome sequence of Drosophila melanogaster";
 Science 287:2185-2195(2000).
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
 CC is thought to mediate the attachment, migration, and organization
 CC of cells into tissues during embryonic development by interacting
 CC with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M95811; AAD19752.1; -
 CC DR EMBL; M19525; AAA28663.1; -
 CC DR EMBL; AE003618; AAF52563.1; -
 CC DR PIR; A28783; WMFB1.
 CC DR HSP; P02488; ITLE.
 CC DR FlyBase; FBgn0002527; LanB1.
 CC DR InterPro; IPR000561; EGF-like.
 CC DR InterPro; IPR001886; LamNT.
 CC DR InterPro; IPR002049; Laminin_EGF.
 CC DR Pfam; PF00053; laminin_EGF; 13.
 CC DR Pfam; PF00055; laminin_Nterm; 1.
 CC DR PRINTS; PRO0011; EGFLAMININ.
 CC DR ProDom; PD002082; LamNT; 1.
 CC DR SMART; SM00180; EGF_Lam; 11.
 CC DR SMART; SM00001; EGF_Like; 1.
 CC DR SMART; SM00136; LamNT; 1.
 CC DR PROSITE; PS00022; EGF_1; 10.
 CC DR PROSITE; PS01186; EGF_2; 2.
 CC DR PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
 CC KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 25 LAMININ BETA-1 CHAIN.
 FT CHAIN 26 1790 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 43 289 LAMININ EGF-LIKE 1.
 FT DOMAIN 290 356 LAMININ EGF-LIKE 2.
 FT DOMAIN 357 419 LAMININ EGF-LIKE 3.
 FT DOMAIN 420 479 LAMININ EGF-LIKE 4.
 FT DOMAIN 480 530 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 531 561 LAMININ DOMAIN IV.
 FT DOMAIN 562 789 LAMININ EGF-LIKE 6.
 FT DOMAIN 791 838 LAMININ EGF-LIKE 7.
 FT DOMAIN 839 884 LAMININ EGF-LIKE 8.
 FT DOMAIN 885 934 LAMININ EGF-LIKE 9.
 FT DOMAIN 935 992 LAMININ EGF-LIKE 10.
 FT DOMAIN 993 1044 LAMININ EGF-LIKE 11.
 FT DOMAIN 1045 1095 LAMININ EGF-LIKE 12.
 FT DOMAIN 1096 1143 LAMININ EGF-LIKE 13.
 FT DOMAIN 1144 1190 DOMAIN II.
 FT DOMAIN 1191 1407 DOMAIN I.
 FT DOMAIN 1408 1434 COILED COIL (POTENTIAL).
 FT DOMAIN 1435 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).
 FT DOMAIN 1455 1507 COILED COIL (POTENTIAL).
 FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).
 FT DOMAIN 1610 1764 BY SIMILARITY.
 FT DISULFID 290 299 BY SIMILARITY.
 FT DISULFID 292 320 BY SIMILARITY.
 FT DISULFID 322 331 BY SIMILARITY.
 FT DISULFID 334 354 BY SIMILARITY.
 FT DISULFID 357 366 BY SIMILARITY.
 FT DISULFID 359 384 BY SIMILARITY.
 FT DISULFID 387 396 BY SIMILARITY.
 FT DISULFID 399 417 BY SIMILARITY.
 FT DISULFID 420 433 BY SIMILARITY.
 FT DISULFID 422 448 BY SIMILARITY.
 FT DISULFID 450 459 BY SIMILARITY.
 FT DISULFID 462 477 BY SIMILARITY.
 FT DISULFID 480 493 BY SIMILARITY.
 FT DISULFID 482 500 BY SIMILARITY.
 FT DISULFID 502 511 BY SIMILARITY.
 FT DISULFID 514 528 BY SIMILARITY.
 FT DISULFID 791 803 BY SIMILARITY.
 FT DISULFID 793 810 BY SIMILARITY.
 FT DISULFID 812 821 BY SIMILARITY.
 FT DISULFID 824 836 BY SIMILARITY.
 FT DISULFID 839 851 BY SIMILARITY.
 FT DISULFID 851 858 BY SIMILARITY.

```
FT DISULFID 860 869 BY SIMILARITY.
FT DISULFID 872 882 BY SIMILARITY.
FT DISULFID 885 894 BY SIMILARITY.
FT DISULFID 887 901 BY SIMILARITY.
FT DISULFID 904 913 BY SIMILARITY.
FT DISULFID 916 932 BY SIMILARITY.
FT DISULFID 935 951 BY SIMILARITY.
FT DISULFID 937 962 BY SIMILARITY.
FT DISULFID 964 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 993 1007 BY SIMILARITY.
FT DISULFID 995 1014 BY SIMILARITY.
FT DISULFID 1017 1026 BY SIMILARITY.
FT DISULFID 1029 1042 BY SIMILARITY.
FT DISULFID 1045 1059 BY SIMILARITY.
FT DISULFID 1047 1066 BY SIMILARITY.
FT DISULFID 1068 1077 BY SIMILARITY.
FT DISULFID 1080 1093 BY SIMILARITY.
FT DISULFID 1096 1108 BY SIMILARITY.
FT DISULFID 1098 1115 BY SIMILARITY.
FT DISULFID 1117 1126 BY SIMILARITY.
FT DISULFID 1129 1141 BY SIMILARITY.
FT DISULFID 1144 1156 BY SIMILARITY.
FT DISULFID 1146 1163 BY SIMILARITY.
FT DISULFID 1165 1174 BY SIMILARITY.
FT DISULFID 1177 1188 BY SIMILARITY.
FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1768 1788 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 643 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.2%; Score 102; DB 1; Length 1790;
Best Local Similarity 21.4%; Pred. No. 0.75;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

Qy 97 GYCD-----CF--ASGDFCNCC-----NCNCC-----CNLLHHDIERFKA 129
Db 1005 GNCDRQTGACLKLYQTGDHCELCCKDGFGLDQONCQCECDPLGNTNNTIAHCDRTG 1064

Qy 130 IKACLRNPEAFQPKIGKGLGNVKNPHK-----GCNRRSGCLKNYCEYEAQIM 181
Db 1065 QCPCL-----PNVQGVRCDOCAENHWKIASGEGCSCNCDPIGALHEQCNSYTGQC 1116

Qy 182 CS---SICKCIGCK-NYEESPERKTLMSMPNMQTGGLEGSHYLPPTKFS--GLPRFSHD 235
Db 1117 CKPGEGGRACQOAHYWNFNPK-----CQPCEDQFGRADFQCD 1157

Qy 236 RPPSCISWEVVEATCCLLAQGEAEKHCCKSLAEQMIILEEFGRLS 284
Db 1158 RETGNCVCHGIGGYKNCACARGYQGFPHCSPC-----GECFN 1196

RESULT 8
TFH2_MOUSE
AC Q9JIB4; STANDARD; PRT; 396 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE TFIIH basal transcription factor complex p44 subunit (Basic
DE transcription factor 2.44 kDa subunit) (BTF2-p44) (General
DE transcription factor IIF polypeptide 2).
GN GTF2H2 OR BTF2P44.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20414747; PubMed=10958627;
RA Endrizzi M.G., Hadinoto V., Gowney J.D., Miller W., Dietrich W.F.;
RL "Genomic sequence analysis of the mouse Naip gene array.";
RL Genome Res. 10:1095-1102(2000).
CC -!- FUNCTION: Component of the core-TFIIH basal transcription factor
CC involved in nucleotide excision repair (NER) of DNA and, when
CC complexed to CAK, in RNA transcription by RNA polymerase II. The
CC N-terminus interacts with and regulates XPD whereas an intact C-
CC terminus is required for a successful escape of RNAP II from the
CC promoter (By similarity).
CC -!- SUBUNIT: One of the six subunits forming the core-TFIIH basal
CC transcription factor. Interacts with XPB, XPD, p82 and p34 (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 C4-TYPE ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF242432; AAR82753.1; .
CC MGI; MGI:1345669; Gtf2h2.
CC InterPro; IPR004595; Ssl1.
CC InterPro; IPR001841; Znf_ring.
CC SMART; SM00327; VWA; 1.
CC TIGRFAMs; TIGR00622; ssl1; 1.
CC PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
CC PROSITE; PS00089; ZF_RING_2; FALSE_NEG.
CC Transcription regulation; DNA repair; Nuclear protein; Zinc-finger.
CC ZN_FING 292 309 C4-TYPE.
CC ZN_FING 346 383 RING-TYPE.
CC DOMAIN 242 245 POLY-SER.
CC SEQUENCE 396 AA; 44687 MW; 1BC57AFD464AA83F CRC64;

Query Match 6.2%; Score 101.5; DB 1; Length 396;
Best Local Similarity 21.8%; Pred. No. 0.16;
Matches 48; Conservative 27; Mismatches 72; Indels 73; Gaps 11;

Qy 2 VICQLKGTQMLCIDNSRTRELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTKL 61
Db 210 VLARETGTYHVLDETHYKELLAAHHVSPPPASSSS-----ECSLRMGFPQHTIA 260

Qy 62 NLITQOLEGALPSVVVNGSAFPGSGTLPGPKITLAGYCDPAGDFCNCCNCC----- 116
Db 261 SLSDQD---AKPS-FSMAHLDNNSTEPG---LTLGGY-----FCPCRAKYCELPVE 305

Qy 117 -----CNLLHH-----DIERPKAKACLRNPEAFQPKIGKGLGN 152
Db 306 CKICGLTLVSAPHILARSYHHLFPLDAFOEISLEEKYGRFCYGCQ-----GE 352

Qy 153 VKPQHNGCN-CRRSGCLKNYCEYEAQIMCSICKICGC 191
Db 353 LKDQHYVYCTVQCNVFCVD--CDVF-----VHDSLHCCFCGC 386

RESULT 9
BAR3_CHITE
ID BAR3_CHITE STANDARD; PRT; 1700 AA.
AC Q03376;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Balbiani ring protein 3 precursor.
GN BR3.
```


FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 3011;
 Best Local Similarity 24.8%; Pred. No. 1.6;
 Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAFF-SGSTLPKPPKTTLAGYDCDFASGDFCNCCNCCNLL-----HHDIERFKAI 130
 DB 1424 SVIPTSGLVVVATDALMTGY-----TGDFDSVDCNTCTVTQVDFSLDPTTFITITLP 1478
 QY 131 KACLGRNPEAFQPKIGKGLG---NVKPOHKGKNCRRSGCLKN--YCECYEAQIMCSSI 185
 DB 1479 QDAVSRTQR--RGRTGRGPGIYRFVAPCE-----RPSGMFSDSSVLCICYDA----- 1523
 QY 186 CKCIGCKNYEESPERRKTLMSMNYAQTGGL-----EG-----SHYLPPTKF 226
 DB 1524 ----GCAMYLELP-AETTIVRLRAYMNTPGLPVCDHLEFWEGVFTGLTHIDAHFLSQTKQ 1578
 QY 227 SG--LPRFSDRRPSSCISWEVEATCACLQAQGEAEKEHCKSL 270
 DB 1579 SGENLP-----YLVAYQATVCARAQAPPSPWDQMWKCL 1611

RESULT 11

EZ_DROME
 ID EZ_DROME STANDARD; PRT; 760 AA.
 AC P42124;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Enhancer of zeste protein.
 DE E(2).
 GN Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94019309; PubMed=8413234;
 RA Jones R.S., Gelbart W.M.;

RT "The Drosophila Polycomb-group gene Enhancer of zeste contains a region with sequence similarity to trithorax.";
 RL Mol. Cell. Biol. 13:6357-6366(1993).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=97164720; PubMed=9012527;
 RA Carrington E.A., Jones R.S.;
 RT "The Drosophila Enhancer of zeste gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution.";
 RL Development 122:4073-4083(1996).
 CC -!- FUNCTION: NEGATIVELY REGULATES THE SEGMENT IDENTITY GENES OF THE ANTENNAPEDIA (ANT-C) AND BITHORAX (BX-C) GENE COMPLEXES.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE EZ FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00180; AAC46462.1; -
 DR FlyBase; FBgn0000629; E(z).
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR InterPro; IPR001214; SET.
 DR Pfam; PF00856; SET; 1.
 DR SMART; SM00395; SANT; 1.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS0280; SET; 1.
 DR Transcription regulation; Nuclear protein; DNA-binding; Repressor;
 KW Developmental protein.
 FT DOMAIN 505 510 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 538 619 CYS-RICH.
 FT DOMAIN 632 743 SET.
 SQ SEQUENCE 760 AA; 86965 MW; 4A28A8EAD7968AC4 CRC64;

Query Match 6.1%; Score 100.5; DB 1; Length 760;
 Best Local Similarity 20.3%; Pred. No. 0.4;
 Matches 36; Conservative 23; Mismatches 63; Indels 55; Gaps 8;

QY 99 CDCFAGDFCN--CNCNCCNLLHHDIERFKAICACLRNPEAFQPKIGKGLGNVQKH 157
 DB 549 CSCIQTONCEFCNCCSDCQN-----RP----- 572
 QY 158 NKGCMRRSGCLKNYCECYEAQIMCS-SICKGCKNYEESPERRKTLMSMNYMOTGGLE 216
 DB 573 -PGCRC-KAQCNTKQPCYLAVERCDPLCQACGADQF-----KLTKITCKNVQVQRLH 625
 QY 217 GSHYLPPTKFSGLPRFSDRRPSSCISWEVEATCACLQAQ-----GEAEKEHCKS 267
 DB 626 KHLMLAPSDIAGWIGIFLKECAQKN-----EFISEYCGEIIISDEADRRGRVYDKYMC 678

RESULT 12
 LMA5_HUMAN
 ID LMA5_HUMAN STANDARD; PRT; 3695 AA.
 AC O15230; Q9H1PI; Q8WZA7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-5 chain precursor.
 GN LAMA5 OR KIAA0533 OR KIAA1907.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Basley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivasalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McTay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
RN SEQUENCE OF 197-1934 FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=21456161; PubMed=11572484;
RX Nagase T., Kikuno R., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
RL DNA Res. 8:179-187(2001).
RN [3]
RN SEQUENCE OF 2051-3695 FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=98290545; PubMed=9628581;
RX Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [4]
RN SEQUENCE OF 2743-3695 FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=97415425; PubMed=9271224;
RA Durkin M.E., Loeschel F., Mattei M.-G., Gilpin B.J., Albrechtsen R.,
RA Wewer U.M.;
RT "Tissue-specific expression of the human laminin alpha5-chain, and
RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal
RT mouse chromosome 2 near the locus for the ragged (Ra) mutation.";
RL FEBS Lett. 411:296-300(1997).
RN [5]
RN EXPRESSION IN RETINA.
RX MEDLINE=20422761; PubMed=10964957;
RA Libby R.T., Champiaud M.-F., Claudepierre T., Xu Y., Gibbons E.P.,
RA Koch M., Burgeson R.E., Hunter D.D., Brunken W.J.;
RT "Laminin expression in adult and developing retinae: evidence of two
RT novel CNS laminins.";
RL J. Neurosci. 20:6517-6528(2000).
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION AND ORGANIZATION
CC OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -1- SUBUNIT: LAMININ-15 COMPLEX IS AN HETEROTRIMER COMPOSED OF THREE
CC CHAINS (ALPHA-5/BETA-2/GAMMA-3) WHICH ARE BOUND TO EACH OTHER BY
CC DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG
CC AND THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).

CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, LUNG, KIDNEY, SKELETAL
CC MUSCLE, PANCREAS, RETINA AND PLACENTA. LITTLE OR NO EXPRESSION IN
CC BRAIN AND LIVER.
CC -1- DOMAIN: DOMAIN G IS GLOBULAR AND IS PART OF THE MAJOR CELL-BINDING
CC SITE LOCATED IN THE LONG ARM OF THE LAMININ HETEROTRIMER.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL354836; CAC22309.1; ALT_SEQ.
CC EMBL; AL354836; CAC22310.1;
CC EMBL; AB067494; BAB67800.1;
CC EMBL; AB011105; BAA25459.1;
CC EMBL; Z95636; CAB09137.1;
CC HSP; P02468; IKLO.
CC Genew; HGNC:6485; LAMA5.
CC MIM; 601033;
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC Pfam; PF00053; laminin_EGF; 2.
CC Pfam; PF00054; laminin_G; 2.
CC ProDom; PD002082; LamNT; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC ProSITE; PS00022; EGF_1; 19.
CC ProSITE; PS01186; EGF_2; 3.
CC ProSITE; PS01248; LAMININ_TYPE_EGF; 19.
CC ProSITE; PS00025; LAM_G_DOMAIN; 5.
CC Glycoproteo; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC -----
CC SIGNAL 35
CC CHAIN 1 35
CC FT CHAIN 36 3695 LAMININ ALPHA-5 CHAIN.
CC FT DOMAIN 36 299 LAMININ N-TERMINAL (DOMAIN VI).
CC FT DOMAIN 300 358 LAMININ EGF-LIKE 1.
CC FT DOMAIN 359 428 LAMININ EGF-LIKE 2.
CC FT DOMAIN 429 474 LAMININ EGF-LIKE 3.
CC FT DOMAIN 494 540 LAMININ EGF-LIKE 4.
CC FT DOMAIN 541 586 LAMININ EGF-LIKE 5.
CC FT DOMAIN 587 631 LAMININ EGF-LIKE 6.
CC FT DOMAIN 632 676 LAMININ EGF-LIKE 7.
CC FT DOMAIN 677 722 LAMININ EGF-LIKE 8.
CC FT DOMAIN 723 775 LAMININ EGF-LIKE 9.
CC FT DOMAIN 776 828 LAMININ EGF-LIKE 10.
CC FT DOMAIN 829 850 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC FT DOMAIN 851 1437 LAMININ DOMAIN IV 1 (DOMAIN IV B).
CC FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.
CC FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.
CC FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.
CC FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.
CC FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC FT DOMAIN 1638 1830 LAMININ DOMAIN IV 2 (DOMAIN IV A).
CC FT DOMAIN 1831 1863 LAMININ EGF-LIKE 16 (C-TERMINAL).
CC FT DOMAIN 1864 1912 LAMININ EGF-LIKE 17.
CC FT DOMAIN 1913 1968 LAMININ EGF-LIKE 18.
CC FT DOMAIN 1969 2022 LAMININ EGF-LIKE 19.
CC FT DOMAIN 2023 2069 LAMININ EGF-LIKE 20.
CC FT DOMAIN 2070 2116 LAMININ EGF-LIKE 21.
CC FT DOMAIN 2117 2166 LAMININ EGF-LIKE 22.
CC FT DOMAIN 2167 2735 DOMAIN II AND I.
CC FT DOMAIN 2736 2929 LAMININ G-LIKE 1.
CC FT DOMAIN 2941 3115 LAMININ G-LIKE 2.
CC FT DOMAIN 3124 3292 LAMININ G-LIKE 3.

```

FT DOMAIN 3340 3513 LAMININ G-LIKE 4.
FT DOMAIN 3520 3692 LAMININ G-LIKE 5.
FT DOMAIN 2221 COILED COIL (POTENTIAL).
FT DOMAIN 2335 2466 COILED COIL (POTENTIAL).
FT DOMAIN 2510 2670 COILED COIL (POTENTIAL).
FT SITE 1722 1724 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1838 1840 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 300 309 BY SIMILARITY.
FT DISULFID 302 322 BY SIMILARITY.
FT DISULFID 324 333 BY SIMILARITY.
FT DISULFID 336 356 BY SIMILARITY.
FT DISULFID 359 368 BY SIMILARITY.
FT DISULFID 361 393 BY SIMILARITY.
FT DISULFID 396 405 BY SIMILARITY.
FT DISULFID 408 426 BY SIMILARITY.
FT DISULFID 429 440 BY SIMILARITY.
FT DISULFID 431 447 BY SIMILARITY.
FT DISULFID 449 458 BY SIMILARITY.
FT DISULFID 461 471 BY SIMILARITY.
FT DISULFID 494 506 BY SIMILARITY.
FT DISULFID 496 515 BY SIMILARITY.
FT DISULFID 517 526 BY SIMILARITY.
FT DISULFID 529 538 BY SIMILARITY.
FT DISULFID 541 553 BY SIMILARITY.
FT DISULFID 543 560 BY SIMILARITY.
FT DISULFID 562 571 BY SIMILARITY.
FT DISULFID 574 584 BY SIMILARITY.
FT DISULFID 587 599 BY SIMILARITY.
FT DISULFID 589 605 BY SIMILARITY.
FT DISULFID 607 616 BY SIMILARITY.
FT DISULFID 619 629 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT DISULFID 634 650 BY SIMILARITY.
FT DISULFID 652 661 BY SIMILARITY.
FT DISULFID 664 674 BY SIMILARITY.
FT DISULFID 677 689 BY SIMILARITY.
FT DISULFID 679 696 BY SIMILARITY.
FT DISULFID 698 707 BY SIMILARITY.
FT DISULFID 710 720 BY SIMILARITY.

Query Match 6.1%; Score 100; DB 1; Length 3695;
Best Local Similarity 21.8%; Pred. No. 2.4;
Matches 57; Conservative 23; Mismatches 100; Indels 82; Gaps 14;

QY 93 ITLAGVCDCAFSGDFC-----NNC--NCNCCNLLHHDIEREKAIRA- 132
DB 294 ISIGRCVCHGADACDAKDPDPFLQCTCOHNTCGGTCDCRCPCFNQ--QFWKRPATAN 351
QY 133 -----CLGR-----NPE-----AFQPKIGKQGLGNV---KPQHNKGCNCRRSGL 169
DB 352 SANECQSCNCGYHATDCYYDPEVDRRRASOSLDGYVGGGVCDICQHHHTGVNCR--CL 409
QY 170 KNYCEYENAIKNSICKICGKN-----YESPERKTLMSPNY-----MOTGGLEG 217
DB 410 PGFYRSPNPLPSPHVCRNCNESDFTDGTCDLTGR--CYCRPNFSGRCDCVCAEGFTG 467
QY 218 SHVLPTRFSGGLPRSHDRRPSSTCSWEVVEATCACLLAQGEAEKEHCKSLAQEMILE 277
DB 468 FPCSTPT-----PSSNDTREQVLFPAGQIVNCDCSAAGTQGNACKRD-----P 510
QY 278 EFGRCLS-----QILHTEFSKG 295
DB 511 RVGRCLKPNFQGTCELCAFG 532

RESULT 13
ID ITB6_MOUSE STANDARD; PRT: 787 AA.
AC Q920T9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin beta-6 precursor.

GN ITGB6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=20547423; PubMed=11095652;
RA Arend L.J., Smart A.M., Briggs J.P.;
RT "Mouse beta(6) integrin sequence, pattern of expression, and role in
RL kidney development.";
RL J. Am. Soc. Nephrol. 11:2297-2305(2000).
CC !- FUNCTION: INTEGRIN ALPHA-V/BETA-6 IS A RECEPTOR FOR FIBRONECTIN
CC AND CYTOTACTIN. IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGANDS
CC (BY SIMILARITY).
CC !- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-6
CC ASSOCIATES WITH ALPHA-V (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC !- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF115376; AAD17212.1;
CC HSSP: P05106; 1JV2.
CC MGD: MGI:96615; Itgb6
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002369; Integrin_B.
CC InterPro: IPR001169; Integrin_beta_C.
CC InterPro: IPR003659; Plexin-like.
CC Pfam: PF00362; Integrin_B; 1.
CC PRINTS: P00186; INTEGRINB.
CC ProDom: PD001811; Integrin_B; 1.
CC SMART: SM00001; EGF_like; 1.
CC SMART: SM00187; INB; 1.
CC SMART: SM00423; PSI; 1.
CC SMART: SM00327; VWFA; 1.
CC PROSITE: PS00243; INTEGRIN_BETA; 2.
CC PROSITE: PS00022; EGF_1; UNKNOWN_2.
CC PROSITE: PS01186; EGF_2; UNKNOWN_1.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 787 INTEGRIN BETA-6.
FT DOMAIN 22 706 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 707 729 POTENTIAL.
FT DOMAIN 730 787 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 131 371 VWFA-LIKE.
FT DOMAIN 456 619 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 456 501 I.
FT REPEAT 502 543 II.
FT REPEAT 544 582 III.
FT REPEAT 583 619 IV.
FT DISULFID 23 454 BY SIMILARITY.
FT DISULFID 31 41 BY SIMILARITY.
FT DISULFID 34 70 BY SIMILARITY.
FT DISULFID 44 59 BY SIMILARITY.
FT DISULFID 197 204 BY SIMILARITY.
FT DISULFID 252 293 BY SIMILARITY.
FT DISULFID 394 406 BY SIMILARITY.
FT DISULFID 426 669 BY SIMILARITY.
FT DISULFID 452 456 BY SIMILARITY.
FT DISULFID 467 479 BY SIMILARITY.
FT DISULFID 476 511 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
```


DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
DR PROSITE; PS00025; LAM_G_DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
FT SIGNAL 1 22
FT CHAIN 23 3106 LAMININ ALPHA-2 CHAIN.
FT DOMAIN 23 3106 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 23 339 LAMININ EGF-LIKE 1.
FT DOMAIN 283 339 LAMININ EGF-LIKE 2.
FT DOMAIN 340 409 LAMININ EGF-LIKE 3.
FT DOMAIN 410 464 LAMININ EGF-LIKE 4.
FT DOMAIN 465 513 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 514 523 LAMININ EGF-LIKE 6.
FT DOMAIN 524 719 LAMININ EGF-LIKE 7.
FT DOMAIN 720 752 LAMININ EGF-LIKE 8.
FT DOMAIN 753 802 LAMININ EGF-LIKE 9.
FT DOMAIN 803 860 LAMININ EGF-LIKE 10.
FT DOMAIN 861 913 LAMININ EGF-LIKE 11.
FT DOMAIN 914 962 LAMININ EGF-LIKE 12.
FT DOMAIN 963 1009 LAMININ EGF-LIKE 13.
FT DOMAIN 1010 1055 LAMININ EGF-LIKE 14.
FT DOMAIN 1056 1101 LAMININ EGF-LIKE 15.
FT DOMAIN 1102 1161 LAMININ EGF-LIKE 16.
FT DOMAIN 1162 1171 LAMININ EGF-LIKE 17.
FT DOMAIN 1172 1375 LAMININ EGF-LIKE 18 (N-TERMINAL).
FT DOMAIN 1376 1415 LAMININ EGF-LIKE 19 (DOMAIN IV A).
FT DOMAIN 1416 1464 LAMININ EGF-LIKE 20 (C-TERMINAL).
FT DOMAIN 1465 1522 LAMININ EGF-LIKE 21.
FT DOMAIN 1523 1569 LAMININ EGF-LIKE 22.
FT DOMAIN 1570 2140 LAMININ II AND I.
FT DOMAIN 2141 2324 LAMININ G-LIKE 1.
FT DOMAIN 2325 2517 LAMININ G-LIKE 2.
FT DOMAIN 2518 2706 LAMININ G-LIKE 3.
FT DOMAIN 2707 2930 LAMININ G-LIKE 4.
FT DOMAIN 2931 3106 LAMININ G-LIKE 5.
FT DOMAIN 3107 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1864 2146 COILED COIL (POTENTIAL).
FT DOMAIN 2147 283 LAMININ G-LIKE 6.
FT DOMAIN 284 292 LAMININ G-LIKE 7.
FT DOMAIN 293 303 LAMININ G-LIKE 8.
FT DOMAIN 304 314 LAMININ G-LIKE 9.
FT DOMAIN 315 337 LAMININ G-LIKE 10.
FT DOMAIN 338 349 LAMININ G-LIKE 11.
FT DOMAIN 350 374 LAMININ G-LIKE 12.
FT DOMAIN 375 386 LAMININ G-LIKE 13.
FT DOMAIN 387 407 LAMININ G-LIKE 14.
FT DOMAIN 408 422 LAMININ G-LIKE 15.
FT DOMAIN 423 438 LAMININ G-LIKE 16.
FT DOMAIN 439 449 LAMININ G-LIKE 17.
FT DOMAIN 450 462 LAMININ G-LIKE 18.
FT DOMAIN 463 478 LAMININ G-LIKE 19.
FT DOMAIN 479 482 LAMININ G-LIKE 20.
FT DOMAIN 483 493 LAMININ G-LIKE 21.
FT DOMAIN 494 511 LAMININ G-LIKE 22.
FT DOMAIN 512 762 LAMININ G-LIKE 23.
FT DOMAIN 763 769 LAMININ G-LIKE 24.
FT DOMAIN 770 781 LAMININ G-LIKE 25.
FT DOMAIN 782 800 LAMININ G-LIKE 26.
FT DOMAIN 801 818 LAMININ G-LIKE 27.
FT DOMAIN 819 828 LAMININ G-LIKE 28.
FT DOMAIN 829 840 LAMININ G-LIKE 29.
FT DOMAIN 841 858 LAMININ G-LIKE 30.
FT DOMAIN 859 875 LAMININ G-LIKE 31.
FT DOMAIN 876 882 LAMININ G-LIKE 32.
FT DOMAIN 883 894 LAMININ G-LIKE 33.
FT DOMAIN 895 911 LAMININ G-LIKE 34.
FT DOMAIN 912 926 LAMININ G-LIKE 35.
FT DOMAIN 927 933 LAMININ G-LIKE 36.
FT DOMAIN 934 944 LAMININ G-LIKE 37.
FT DOMAIN 945 960 LAMININ G-LIKE 38.
FT DOMAIN 961 975 LAMININ G-LIKE 39.
FT DOMAIN 976 981 LAMININ G-LIKE 40.
FT DOMAIN 982 992 LAMININ G-LIKE 41.
FT DOMAIN 993 1007 LAMININ G-LIKE 42.
FT DOMAIN 1008 1019 LAMININ G-LIKE 43.

FT DISULFID 1012 1026 BY SIMILARITY.
FT DISULFID 1028 1037 BY SIMILARITY.
FT DISULFID 1040 1053 BY SIMILARITY.
FT DISULFID 1056 1068 BY SIMILARITY.
FT DISULFID 1077 1086 BY SIMILARITY.
FT DISULFID 1089 1099 BY SIMILARITY.
FT DISULFID 1416 1425 BY SIMILARITY.
FT DISULFID 1418 1432 BY SIMILARITY.
FT DISULFID 1435 1444 BY SIMILARITY.
FT DISULFID 1447 1462 BY SIMILARITY.
FT DISULFID 1465 1480 BY SIMILARITY.
FT DISULFID 1467 1490 BY SIMILARITY.
FT DISULFID 1493 1502 BY SIMILARITY.
FT DISULFID 1505 1520 BY SIMILARITY.
FT DISULFID 1523 1535 BY SIMILARITY.
FT DISULFID 1525 1542 BY SIMILARITY.
FT DISULFID 1544 1553 BY SIMILARITY.
FT DISULFID 1556 1567 BY SIMILARITY.
FT DISULFID 1570 1570 INTERCHAIN (PROBABLE).
FT DISULFID 1574 1574 INTERCHAIN (PROBABLE).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.0%; Score 98.5; DB 1; Length 3106;
Best Local Similarity 22.0%; Pred. No. 2.7;
Matches 70; Conservative 21; Mismatches 82; Indels 145; Gaps 19;

Qy 73 PSVNGSAPP-----SGSTL---PGPKITLAGYCDGFAS---GDFFC 108
Db 851 PSVGGSCQPCQNDNLDYSPGSCDSLSGCLCKPGN---TTRYCELCDAGYFGDAV 906
Qy 109 NNCNCCN---NLHHDIERFAKRAKCLGRN-----PEAFQPKIGKQGLGNVNP 155
Db 907 NTKNCPQCRDINGSFSDCHTRTQCECRPNVQGRHCDCKPETFGLQGRGCL---- 961
Qy 156 QHNKGCNRRSGCLKNYCEYEAQIMCS---SICKICGCK-----NYEE----- 196
Db 962 ---PCNCSFGSKSFDCEA---SGQCWCPGVAGKKCDRAHGYPNFQEGGCIACDCSHLG 1016
Qy 197 ---SPERTKLSMPNMYMOTGLEGSHYLPPT----- 224
Db 1017 NNCDPRTQCICPPN---TTGEKCECLPNTWGHISIVTGCKVCNCSTVGSLSAQCNVNTG 1073
Qy 225 -----KFSGLPRFSDRRPSSCI---SWEV-----VEAT-----CACLLA 256
Db 1074 QCSCHPKFSGM-----KCSECSRGHWNYPCLTCLDCFLPCTDATTCDLETRKCSQDQ 1126
Qy 257 QGE-----EAEKEHCKSC 269
Db 1127 TQCCKVNVGVGHCDRC 1144

RESULT 15
ZAN_MOUSE
ID ZAN_MOUSE STANDARD; PRT; 5376 AA.
AC O88799; O08647;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;

RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT "Species diversity in the structure of zonadhesin, a sperm-specific
FT membrane protein containing multiple cell adhesion molecule-like
FT domains";
RL J. Biol. Chem. 273:3415-3421(1998).
RN [2]
RP SEQUENCE OF 4864-5376 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
FT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD.
CC -1- TISSUE SPECIFICITY: IN TESTIS: PRIMARILY IN HAPLOID SPERMATIDS.
CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -1- DOMAIN: THE WVD DOMAIN 2 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
CC -1- SIMILARITY: CONTAINS 25 WVD DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U97068; AAC26680.1; -
DR EMBL: U83190; AAC53125.1; -
DR MGD; MGI:108656; Zeb.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR000998; MAM_domain.
DR InterPro: IPR002919; TIL_Cysrich.
DR InterPro: IPR003328; TIL_Cysrich.
DR InterPro: IPR001007; WVF_C.
DR InterPro: IPR001846; WVF_D.
DR Pfam; PF00094; wvd; 4.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILa; 25.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FOLN; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; WVC; 17.
DR SMART; SM00216; WVD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS50060; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 5376 ZONADHESIN.
FT DOMAIN 18 5310. EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5311 5337 POTENTIAL.
FT DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 45 210 MAM 1.
FT DOMAIN 215 374 MAM 2.
FT DOMAIN 377 542 MAM 3.

FT	DOMAIN	547	1170	80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT	DOMAIN	1171	1280	(MUCIN-LIKE DOMAIN).
FT	DOMAIN	1281	1669	WVFD 1 (PARTIAL).
FT	DOMAIN	1670	2056	WVFD 2.
FT	DOMAIN	2057	2459	WVFD 3.
FT	DOMAIN	2460	2579	WVFD 4.
FT	DOMAIN	2580	2699	WVFD 5 (PARTIAL).
FT	DOMAIN	2700	2819	WVFD 6 (PARTIAL).
FT	DOMAIN	2820	2939	WVFD 7 (PARTIAL).
FT	DOMAIN	2940	3059	WVFD 8 (PARTIAL).
FT	DOMAIN	3060	3179	WVFD 9 (PARTIAL).
FT	DOMAIN	3180	3299	WVFD 10 (PARTIAL).
FT	DOMAIN	3300	3416	WVFD 11 (PARTIAL).
FT	DOMAIN	3417	3536	WVFD 12 (PARTIAL).
FT	DOMAIN	3537	3656	WVFD 13 (PARTIAL).
FT	DOMAIN	3657	3776	WVFD 14 (PARTIAL).
FT	DOMAIN	3777	3892	WVFD 15 (PARTIAL).
FT	DOMAIN	3893	4029	WVFD 16 (PARTIAL).
FT	DOMAIN	4030	4148	WVFD 17 (PARTIAL).
FT	DOMAIN	4149	4263	WVFD 18 (PARTIAL).
FT	DOMAIN	4264	4383	WVFD 19 (PARTIAL).
FT	DOMAIN	4384	4503	WVFD 20 (PARTIAL).
FT	DOMAIN	4504	4623	WVFD 21 (PARTIAL).
FT	DOMAIN	4624	4743	WVFD 22 (PARTIAL).
FT	DOMAIN	4744	4863	WVFD 23 (PARTIAL).
FT	DOMAIN	4864	5261	WVFD 24 (PARTIAL).
FT	DOMAIN	5259	5295	WVFD 25.
FT	DISULFID	5263	5274	EGF-LIKE.
FT	DISULFID	5268	5283	BY SIMILARITY.
FT	DISULFID	5285	5294	BY SIMILARITY.
FT	CARBOHYD	339	339	BY SIMILARITY.
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1216	1216	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1314	1314	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1814	1814	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1908	1908	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1933	1933	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2028	2028	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2111	2111	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2142	2142	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2332	2332	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2575	2575	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2692	2692	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2812	2812	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3065	3065	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3144	3144	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3172	3172	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3288	3288	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3292	3292	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3782	3782	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4005	4005	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4136	4136	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4243	4243	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4254	4254	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4335	4335	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4376	4376	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4586	4586	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	5136	5136	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	5252	5252	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	5376 AA;	579908 MW;	0B44DB77DF2A2620 CRC64;

Query Match 6.0%; Score 98.5; DB 1; Length 5376;

Best Local Similarity 23.2%; Pred. No. 4.8; Matches 53; Conservative 19; Mismatches 73; Indels 83; Gaps 15;

QY 99 CDCFASGDFCNCCNC-----NCCNNLHDIERFAKKAACL-GRNPEAFQPKIGKQL 150

DB 4232 CACVGAQCCHNFTCTGTGTCQNSCSKITVQCFAHSQYTTCLPCLPSCFDP---EGLC 4288

[illegible]

Search completed: April 21, 2003, 11:49:02
Job time : 15.0606 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:38:48 ; Search time 8.93939 seconds
(without alignments)
1368.719 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGGAGMLCIDNCGA.....GRLSQLHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1599	100.0	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
2	1251	78.2	299	1 MTL5_HUMAN	Q9y415 homo sapien
3	110.5	6.9	971	1 RECK_HUMAN	Q95980 homo sapien
4	103.5	6.5	643	1 CD93_RAT	Q9et61 rattus norv
5	96.5	6.0	1790	1 LMB1_DROME	P11046 drosophila
6	96	6.0	773	1 MSL2_DROME	P50534 drosophila
7	96	6.0	3718	1 LMA5_MOUSE	Q61001 mus musculus
8	95	5.9	1096	1 KPC1_ASPNG	Q00078 aspergillus
9	93.5	5.8	652	1 CD93_HUMAN	Q9npy3 homo sapien
10	93	5.8	1816	1 LMA4_HUMAN	Q16363 homo sapien
11	92	5.8	798	1 ITB1_XENLA	P12606 xenopus lae
12	91.5	5.7	644	1 CD93_MOUSE	Q89103 mus musculus
13	91	5.7	493	1 OC90_HUMAN	Q02509 homo sapien
14	91	5.7	747	1 EZH1_HUMAN	Q92800 homo sapien
15	91	5.7	2907	1 FBN2_MOUSE	Q61555 mus musculus
16	89.5	5.6	459	1 RN25_HUMAN	Q96bh1 homo sapien
17	89	5.6	1786	1 LMB1_MOUSE	P02469 mus musculus
18	88	5.5	798	1 ITB0_XENLA	P12607 xenopus lae
19	88	5.5	1786	1 LMB1_HUMAN	P07942 homo sapien
20	87.5	5.5	971	1 RECK_MOUSE	Q920j1 mus musculus
21	87	5.4	290	1 Y070_NPVAC	P41470 autographa
22	87	5.4	501	1 TRA2_HUMAN	Q12933 homo sapien
23	87	5.4	652	1 SYG_SCHPO	Q10179 schizosacch
24	86.5	5.4	611	1 LEW2_CANFA	P33730 canis famil
25	86	5.4	610	1 LEW2_HUMAN	P16581 homo sapien
26	86	5.4	747	1 EZH1_MOUSE	P70351 mus musculus
27	86	5.4	830	1 LEM3_HUMAN	P16109 homo sapien
28	85.5	5.3	769	1 ITB2_HUMAN	P05107 homo sapien
29	85.5	5.3	806	1 ITB7_MOUSE	P26011 mus musculus
30	85	5.3	398	1 MUB1_XENLA	P38565 xenopus lae
31	85	5.3	746	1 EZH2_MOUSE	Q61188 mus musculus
32	85	5.3	787	1 ITB6_MOUSE	Q920t9 mus musculus
33	84.5	5.3	169	1 KRUA_HUMAN	P26371 homo sapien

34 84.5 5.3 720 1 NRG3_HUMAN P56975 homo sapien
35 84.5 5.3 1639 1 LMG1_DROME P15215 drosophila
36 84.5 5.3 1816 1 LMA4_MOUSE P97927 mus musculus
37 84 5.3 769 1 ITB2_BOVIN P32592 bos taurus
38 84 5.3 2531 1 NTC1_MOUSE Q01705 mus musculus
39 84 5.3 2531 1 NTC1_RAT Q07008 rattus norv
40 84 5.3 3707 1 PGBM_MOUSE Q05793 mus musculus
41 84 5.3 3828 1 TRX_DROVI Q24742 drosophila
42 83.5 5.2 419 1 VEGC_HUMAN P49767 homo sapien
43 83.5 5.2 769 1 ITB2_PIG P53714 sus scrofa
44 83.5 5.2 798 1 ITB7_HUMAN P26010 homo sapien
45 83.5 5.2 2766 1 THYG_MOUSE O08710 mus musculus

ALIGNMENTS

RESULT 1
MTL5_MOUSE
ID MTL5_MOUSE STANDARD; PRT; 295 AA.
AC Q9wtj6:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR, TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
CC PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright.. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U77383; AAD24667.1; -;
DR EMBL: U67176; AAD24666.1; -;
DR MGD: MGI:1340029; Mt15.
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 1.
KW Spermatogenesis.
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Query Match 100.0%; Score 1599; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.8e-131; Indels 0; Gaps 0;
Matches 295; Conservative 0; Mismatches 0;

QY 1 MVICQLKGGAGMLCIDNCGARELKALHLLPQYDDQSSFFQSQSELPKPMPTTLVGRLLPYPAK 60
|||||
Db 1 MVICQLKGGAGMLCIDNCGARELKALHLLPQYDDQSSFFQSQSELPKPMPTTLVGRLLPYPAK 60
|||||
QY 61 LNLITQVDNGALPSAVNGAAPPSPGALQGPPIKTLTSGYCDCFSSGDFCNSCSCNNLRHEL 120
|||||
Db 61 LNLITQVDNGALPSAVNGAAPPSPGALQGPPIKTLTSGYCDCFSSGDFCNSCSCNNLRHEL 120
|||||

Qy 121 ERFKAIAKCLDRNPAFQPMKGRGLGAALRHSGKNCGRSGCLKNKYCEYAKIMCSS 180
 Db 121 ERFKAIAKCLDRNPAFQPMKGRGLGAALRHSGKNCGRSGCLKNKYCEYAKIMCSS 180
 Qy 181 IKCTIACKNYEESPRKMLMSTPHYMEPGDPESSHLYSPAFSGPPKLRKRNQAFSCISW 240
 Db 181 IKCTIACKNYEESPRKMLMSTPHYMEPGDPESSHLYSPAFSGPPKLRKRNQAFSCISW 240
 Qy 241 EWVETACCLAQGEAEQHCSPSLAEQMILEEFGRCLSOILHTEPKSKGLKIE 295
 Db 241 EWVETACCLAQGEAEQHCSPSLAEQMILEEFGRCLSOILHTEPKSKGLKIE 295

RESULT 2

MTL5_HUMAN STANDARD; PRT; 299 AA.
 AC Q9Y4I5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific
 DE metallothionein-like protein).
 GN MTL5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=99208669; PubMed=10191092;
 RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
 RT "A novel testis-specific metallothionein-like protein, tesmin, is an
 RT early marker of male germ cell differentiation.";
 RL Genomics 57:130-136(1999).
 CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U86074; AD24668.1; --
 DR Genew; HGNC:7446; MTL5.
 DR MIM; 604374; --
 DR InterPro; IPR005172; CXC.
 DR Pfam; PF03638; CXC; 1.
 DR Spermatoogenesis.
 FT DOMAIN 99 191 CYS-RICH.
 SO SEQUENCE 299 AA; 33110 MW; 8f679e140f969c11 CRC64;
 Query Match 78.2%; Score 1251; DB 1; Length 299;
 Best Local Similarity 76.3%; Pred. No. 2.6e-101;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;
 Qy 1 MVICQLKGGAGMLCINDCGARELKHLLPQYDQSSFPQSELPKPMTLVGRLLPVPAPK 60
 Db 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDONNYLQSDVPKPMALVGRPLPASTK 60
 Qy 61 LNLITQVNGALPSAVNGAAPPSGALOGPPKPYLISGYCDGFCFSGDCNCSG----CNLL 116
 Db 61 LNLITQLEGALPSVNGSAFSPGSLTLPFGPKPYLAGYCDGFCFSGDCNCCNCCNLL 120
 Qy 117 RHELEFKAIAKACIDRNPAPQPMKGRGLGAALRHSGKNCGRSGCLKNKYCEYAKI 176
 Db 121 HHDIERFKAIAKACLRNPAPQPMKGRGLGAALRHSGKNCGRSGCLKNKYCEYAKI 180
 Qy 177 MCSICKICKACKNYEESPERKMLMSTPHYMEPGDPESSHLYSPAFSGPPKLRKRNQAFS 236

Db 181 MCSICKICKACKNYEESPERKMLMSTPHYMEPGDPESSHLYSPAFSGPPKLRKRNQAFS 240
 Qy 237 CISWEVVEATCACLQAQGEAEQHCSPSLAEQMILEEFGRCLSOILHTEPKSKGLKIE 295
 Db 241 CISWEVVEATCACLQAQGEAEQHCSPSLAEQMILEEFGRCLSOILHTEPKSKGLKIE 299
 RESULT 3
 RECK_HUMAN STANDARD; PRT; 971 AA.
 AC Q95980; Q8WX37;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Reversion-inducing cysteine-rich protein with Kazal motifs precursor
 DE (RRECK) (Suppressor of tumorigenicity 15) (ST15).
 GN RECK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND N-GLYCOSYLATION.
 RC TISSUE=Fibroblast;
 RX MEDLINE=99007295; PubMed=9789069;
 RA Takahashi C., Sheng Z., Horan T.P., Kitayama H., Maki M., Hitomi K.,
 RA Kitaura Y., Takai S., Sasahara R.M., Horimoto A., Ikawa Y.,
 RA Ratzkin B.J., Arakawa T., Noda M.;
 RT "Regulation of matrix metalloproteinase-9 and inhibition of tumor
 RT invasion by the membrane-anchored glycoprotein RECK.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13221-13226(1998).
 RN [2]
 RP SEQUENCE OF 363-971 FROM N.A.
 RA Kimberley A.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Negatively regulates matrix metalloproteinase-9 (MMP-9)
 CC by suppressing MMP-9 secretion and by direct inhibition of its
 CC enzymatic activity. RECK down-regulation by oncogenic signals may
 CC facilitate tumor invasion and metastasis. Appears to also
 CC regulate MMP-2 and MT1-MMP, which are involved in cancer
 CC progression.
 CC -1- SUBUNIT: Interacts with MMP-9.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: Expressed in various tissues and untransformed
 CC cells. It is undetectable in tumor-derived cell lines and
 CC oncogenically transformed cells.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D50406; BAA34060.1; --
 DR EMBL; AL158830; CADI3384.1; --
 DR Genew; HGNC:11345; RECK.
 DR MIM; 605227; --
 DR HSSP; P80424; IAN1.
 DR InterPro; IPR002350; kazal.
 DR SMART; SM00280; KAZAL; 3.
 DR SMART; SM00011; VWC_def; 1.
 DR PROSITE; PS00282; KAZAL; 1.
 DR Signal; Glycoprotein; GPI-anchor; Serine protease inhibitor;
 DR Membrane; Anti-oncogene; Repeat.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 942 REVERSION-INDUCING CYSTEINE-RICH PROTEIN
 FT WITH KAZAL MOTIFS.
 FT PROPEP 943 971 REMOVED IN MATURE FORM (POTENTIAL).
 FT LIPID 942 942 GPI-ANCHOR (POTENTIAL).


```

FT DISULFID 433 446 BY SIMILARITY.
FT DISULFID 448 461 BY SIMILARITY.
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 417 417 E -> K (IN REF. 2).
SQ SEQUENCE 643 AA; 68781 MW; 9AE4C933AD943DB6 CRC64;

Query Match 6.5%; Score 103.5; DB 1; Length 643;
Best Local Similarity 29.4%; Pred. No. 0.16;
Matches 37; Conservative 14; Mismatches 54; Indels 21; Gaps 6;

QY 83 SGPALQGPPTKLTLS---GYC--DCFSGGDFCNSCNLRLHLELRFKAICACLDNRPEAF 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 250 SGP-LCVSPKFCGFSNNGGQQDCFEFGGSPRCG---RPGFRLLDLDVTCASRNPCCSS 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 138 OPKMGKGRGLAAGKLRHSGCKRSGCLKNYCEKVEAKIMSSICKCTACKNYEESPERK 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 NPCTGGGMCHSVPLSENTHCHCPRGYQLDS-----SQVHCVDIDEC-----EDSPCDQ 353
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 198 MLMSTP 203
   :|||
DB 354 ECINTP 359

RESULT 5
LMB1_DROME STANDARD; PRT: 1790 AA.
AC P11046; Q26328; Q9X2T4; Q9V1W6;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LAMB1 OR LAMB1 OR CG7123.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=94000382; PubMed=8397815;
RA Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;
RT "Analysis of the Drosophila gene for the laminin B1 chain.";
RL DNA Cell Biol. 12:573-587(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88210471; PubMed=3365769;
RA Montell D.J., Goodman C.S.;
RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain
   reveals domains of homology with mouse.";
RL Cell 53:463-473(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beson K.V., Benos P.V., Berland B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```

```

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M95811; AAD19752.1; -
CC EMBL; M19525; AAA28663.1; -
CC EMBL; AE003618; AAF52563.1; -
CC PIR; A28783; MMFFB1.
CC HSSP; P02468; 1TLE.
CC Flybase; FBgn0002527; Lmb1.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR002049; Laminin_EGF.
CC Pfam; PF00053; laminin_EGF; 13.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGF_LAMININ.
CC ProDom; PD002082; LamNT; 1.
CC SMART; SM00180; EGF_Lam; 11.
CC SMART; SM00001; EGF-like; 1.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 10.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL 1 25 LAMININ BETA-1 CHAIN.
CC CHAIN 26 1790 LAMININ N-TERMINAL (DOMAIN VI).
CC DOMAIN 43 289 LAMININ EGF-LIKE 1.
CC DOMAIN 290 356

```

```
FT DOMAIN 357 419 LAMININ EGF-LIKE 2.
FT DOMAIN 420 479 LAMININ EGF-LIKE 3.
FT DOMAIN 480 530 LAMININ EGF-LIKE 4.
FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
FT DOMAIN 562 789 LAMININ DOMAIN IV.
FT DOMAIN 791 838 LAMININ EGF-LIKE 6.
FT DOMAIN 839 884 LAMININ EGF-LIKE 7.
FT DOMAIN 885 934 LAMININ EGF-LIKE 8.
FT DOMAIN 935 992 LAMININ EGF-LIKE 9.
FT DOMAIN 993 1044 LAMININ EGF-LIKE 10.
FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.
FT DOMAIN 1096 1143 LAMININ EGF-LIKE 12.
FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
FT DOMAIN 1191 1407 DOMAIN II.
FT DOMAIN 1408 1434 DOMAIN ALPHA.
FT DOMAIN 1435 1790 DOMAIN I.
FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).
FT DOMAIN 1435 1507 COILED COIL (POTENTIAL).
FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).
FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).
FT DISULFID 290 299 BY SIMILARITY.
FT DISULFID 292 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 334 354 BY SIMILARITY.
FT DISULFID 357 366 BY SIMILARITY.
FT DISULFID 359 384 BY SIMILARITY.
FT DISULFID 387 396 BY SIMILARITY.
FT DISULFID 399 417 BY SIMILARITY.
FT DISULFID 420 433 BY SIMILARITY.
FT DISULFID 422 448 BY SIMILARITY.
FT DISULFID 450 459 BY SIMILARITY.
FT DISULFID 462 477 BY SIMILARITY.
FT DISULFID 480 493 BY SIMILARITY.
FT DISULFID 482 500 BY SIMILARITY.
FT DISULFID 502 511 BY SIMILARITY.
FT DISULFID 514 528 BY SIMILARITY.
FT DISULFID 791 803 BY SIMILARITY.
FT DISULFID 793 810 BY SIMILARITY.
FT DISULFID 812 821 BY SIMILARITY.
FT DISULFID 824 836 BY SIMILARITY.
FT DISULFID 839 851 BY SIMILARITY.
FT DISULFID 841 858 BY SIMILARITY.
FT DISULFID 860 869 BY SIMILARITY.
FT DISULFID 872 882 BY SIMILARITY.
FT DISULFID 885 894 BY SIMILARITY.
FT DISULFID 887 901 BY SIMILARITY.
FT DISULFID 904 913 BY SIMILARITY.
FT DISULFID 916 932 BY SIMILARITY.
FT DISULFID 935 951 BY SIMILARITY.
FT DISULFID 937 962 BY SIMILARITY.
FT DISULFID 964 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 993 1007 BY SIMILARITY.
FT DISULFID 995 1014 BY SIMILARITY.
FT DISULFID 1017 1026 BY SIMILARITY.
FT DISULFID 1029 1042 BY SIMILARITY.
FT DISULFID 1045 1059 BY SIMILARITY.
FT DISULFID 1047 1056 BY SIMILARITY.
FT DISULFID 1068 1077 BY SIMILARITY.
FT DISULFID 1080 1093 BY SIMILARITY.
FT DISULFID 1096 1108 BY SIMILARITY.
FT DISULFID 1098 1115 BY SIMILARITY.
FT DISULFID 1117 1126 BY SIMILARITY.
FT DISULFID 1129 1141 BY SIMILARITY.
FT DISULFID 1144 1156 BY SIMILARITY.
FT DISULFID 1146 1163 BY SIMILARITY.
FT DISULFID 1165 1174 BY SIMILARITY.
FT DISULFID 1177 1188 BY SIMILARITY.
FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1788 1798 INTERCHAIN (PROBABLE).
FT SITE 643 645 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC...) (POTENTIAL).
```

Query Match 6.0%; Score 96.5; DB 1; Length 1790;
Best Local Similarity 21.0%; Pred. No. 2;
Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;

QY 94 TLSGYCDYCFSSGDF-----CNCSGSC-----NNLRHELEFRKRAKLRNDEAFQPKM 141
DB 1021 TTGDHCLCKDGFPGDALQNCQCCDFLGTTNTTAHCDRFTGQCPCPCL-----PNV 1072
QY 142 GKRLGAALKLRHSK-----GCNKRSGCLNKYCEYAKIMCSSICIKACK-----N 189
DB 1073 QGVRCDCQCAENHWKIASGEGCESCNDPICALHEQCNSYTGQCCKPFGGGRACNQCQAH 1132
QY 190 YEESPERKMLMSTPHYMEPCDFESSHYLSPAKFSGPPKLRKNQAFSCISWEVVEATCAC 249
DB 1133 YWGNPNKQPCPCDQFGAADFQC-----DRETGNCVCHGEGGYKCN 1175
QY 250 LLAQGEAEQEHCS 264
DB 1176 ECARGYIGQPHCSP 1190

RESULT 6
MSL2_DROME STANDARD; PRT; 773 AA.
AC P50534;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Male-specific lethal-2 protein.
GN MSL-2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95300219; PubMed=7781064;
RA Kuroda M.I.;
RT "Expression of msl-2 causes assembly of dosage compensation
regulators on the X chromosomes and female lethality in Drosophila."
RL Cell 81:867-877(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95317307; PubMed=7796814;
RA Zhou S., Yang Y., Scott M.J., Pannuti A., Fehr K.C., Eisen A.,
Koonin E.V., Fouts D.L., Wrightsman R., Manning J.E., Lucchesi J.C.;
RT "Male-specific lethal 2, a dosage compensation gene of Drosophila,
undergoes sex-specific regulation and encodes a protein with a RING
finger and a metallothionein-like cysteine cluster."
RL EMBO J. 14:2884-2895(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017637; PubMed=7588059;
RA Bashaw G.J., Baker B.S.;
RT "The msl-2 dosage compensation gene of Drosophila encodes a putative
DNA-binding protein whose expression is sex specifically regulated
by sex-lethal."
RL Development 121:3245-3258(1995).
CC -!- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
DOSE COMPENSATION). MSL-2 IS REQUIRED FOR TRANSLATION AND/OR
STABILITY OF MSL-1 IN MALES. BOTH MSL-1 AND MSL-2 BIND TO HUNDREDS
OF SITES ALONG THE MALE X CHROMOSOME, BUT NOT THE FEMALE X OR THE

RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasiaho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.W.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RL "The DNA sequence and comparative analysis of human chromosome 20.";
RN Nature 414:865-871(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP CHARACTERIZATION.
RX PubMed=11994479;
RA McGreal E.P., Ikewaki N., Akatsu H., Morgan B.P., Gasque P.;
RT "Human ClqRP is identical with Cd93 and the mni-11 antigen but does
RT not bind Clq.";
RL J. Immunol. 168:5222-5232(2002).
RN [6]
RN O-GLYCOSYLATION.
RX PubMed=10092817;
RA Nepomuceno R.R., Ruiz S., Park M., Tenner A.J.;
RT "ClqRP is a heavily O-glycosylated cell surface protein involved in
RT the regulation of phagocytic activity.";
RL J. Immunol. 162:3583-3589(1999).
CC -|- FUNCTION: Receptor (or element of a larger receptor complex) for
CC Clq, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SPA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- TISSUE SPECIFICITY: Highly expressed in endothelial cells,
CC platelets, cells of myeloid origin, such as monocytes and
CC neutrophils. Not expressed in cells of lymphoid origin.
CC -|- PTM: N- and O-glycosylated.
CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -|- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -|- CAUTION: Has been sometimes referred to as a collectin receptor.
CC -|- DATABASE: NAME=PROW; NOTE=PROW 3.1-6(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/467246456.g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U94333; AAB53110.1; .
CC EMBL; AL118508; CAC00597.1; .
CC EMBL; BC028075; AAH28075.1; .
CC HSP; P35355; 1EMN.
CC MIM: 120577; .
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR000361; EGF-like.
CC InterPro; IPR001881; EGF-Ca.
CC InterPro; IPR001304; Lectin-C.
CC InterPro; IPR001187; Tissue_factor.
CC Pfam; PF00008; EGF; 5.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF01108; Tissue_fac; 1.
CC SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_CA; 3.
DR SMART; SM00001; EGF_like; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00615; C-TYPE LECTIN_1; FALSE_NEG.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
KW Receptor; EGF-like domain; Signal; Transmembrane; Glycoprotein;
KW Repeat; Lectin; Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 652
FT DOMAIN 24 580
FT TRANSMEM 581 601
FT DOMAIN 602 652
FT DOMAIN 32 174
FT DOMAIN 260 301
FT DOMAIN 302 344
FT DOMAIN 345 384
FT DOMAIN 385 426
FT DOMAIN 427 468
FT DOMAIN 594 601
FT DISULFID 264 275
FT DISULFID 271 285
FT DISULFID 287 300
FT DISULFID 306 317
FT DISULFID 311 328
FT DISULFID 330 343
FT DISULFID 349 358
FT DISULFID 354 367
FT DISULFID 369 383
FT DISULFID 389 400
FT DISULFID 396 409
FT DISULFID 411 425
FT DISULFID 431 443
FT DISULFID 439 452
FT DISULFID 454 467
FT CARBOHYD 325 325
FT VARIANT 318 318
FT CONFLICT 22 22
FT CONFLICT 36 36
FT CONFLICT 38 39
FT CONFLICT 155 155
FT CONFLICT 186 186
FT CONFLICT 492 492
FT CONFLICT 496 496
FT CONFLICT 504 504
FT CONFLICT 541 541
SQ SEQUENCE 652 AA; 68560 MW; EECA0FEAC55FCAC2 CRC64;
Query Match 5.8%; Score 93.5; DB 1; Length 652;
Best Local Similarity 27.0%; Pred. No. 1.2;
Matches 30; Conservative 11; Mismatches 51; Indels 19; Gaps 4;
QY 96 SGYC--DCFSQDFNCSCNNLRHELEFRKAIACLDNRPEAFQPKMGKGLGAALKLRH 153
DB 268 NGGCHQDCEGGDGFGLCC---RPGFRLLDVLVCASRNPCSSSPCRG----- 313
QY 154 SKGCKNCRSGCLUKNY-CECYEAKIMCSSTCKCIACKNYEESPERKMLMSTP 203
DB 314 --GATCVLGPCHKVNTCRCPQGYQLDSSQLDCVDVDECDSPCAQBCVNTP 362
RESULT 10
ID LMA4_HUMAN STANDARD; PRT; 1816 AA.
AC Q16363; Q15335; Q14735; Q9UE18; Q9UJN9;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-4 chain precursor.
GN LAMA4.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal lung;
 RA MEDLINE=9300971; PubMed=7781776;
 RX Iivanainen A., Sainio K., Sariola H., Tryggvason K.;
 RT "Primary structure and expression of a novel human laminin alpha 4
 chain.";
 RL FEBS Lett. 365:183-188(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=97454279; PubMed=9310354;
 RA Richards A.J., Luccarini C., Pope F.M.;
 RT "The structural organisation of LAMA4, the gene encoding laminin
 alpha4.";
 RL Eur. J. Biochem. 248:15-23(1997).
 RN [3]
 RP SEQUENCE OF 236-1816 FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=95048381; PubMed=7959779;
 RA Richards A.J., Al-Imara L., Carter N.P., Lloyd J.C., Leversha M.A.,
 Pope F.M.;
 RT "Localization of the gene (LAMA4) to chromosome 6q21 and isolation of
 a partial cDNA encoding a variant laminin A chain.";
 RL Genomics 22:237-239(1994).
 RN [4]
 RP SEQUENCE OF 66-1816 FROM N.A.
 RA Submy B.;
 RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 different polypeptide chains (alpha, beta, gamma), which are bound
 to each other by disulfide bonds into a cross-shaped molecule
 comprising one long and three short arms with globules at each
 end
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG,
 OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO
 EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE,
 BRAIN. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN
 FETAL AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS
 IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- SIMILARITY: CONTAINS 3.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -1- CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S78569; AAB34635.1; -;
 DR EMBL; X91171; CAA62596.1; -;
 DR EMBL; Y14240; CAA74636.1; -;
 DR EMBL; X76939; CAA54258.1; -;
 DR EMBL; Z99289; CAB16553.1; -;
 DR HSSP; P02468; IKLO.
 DR Genev; HGNC:6484; LAMA4.
 DR MIM; 600133; -;
 DR InterPro; IPR000561; EGF-like.

DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR Pfam; PF00053; laminin_EGF; 6.
 DR Pfam; PF00054; laminin_G; 6.
 DR SMART; SM00180; EGF_Lam; 3.
 DR SMART; SM00282; LamG; 5.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 3.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 1816 LAMININ ALPHA-4 CHAIN.
 FT DOMAIN 82 131 LAMININ EGF-LIKE 1.
 FT DOMAIN 132 186 LAMININ EGF-LIKE 2.
 FT DOMAIN 187 240 LAMININ EGF-LIKE 3.
 FT DOMAIN 241 255 LAMININ EGF-LIKE 4 (INCOMPLETE).
 FT DOMAIN 256 825 DOMAIN II AND I.
 FT DOMAIN 826 1028 LAMININ G-LIKE 1.
 FT DOMAIN 1040 1220 LAMININ G-LIKE 2.
 FT DOMAIN 1227 1395 LAMININ G-LIKE 3.
 FT DOMAIN 1462 1633 LAMININ G-LIKE 4.
 FT DOMAIN 1640 1813 LAMININ G-LIKE 5.
 FT DOMAIN 313 396 COILED COIL (POTENTIAL).
 FT DOMAIN 466 521 COILED COIL (POTENTIAL).
 FT DOMAIN 574 607 COILED COIL (POTENTIAL).
 FT DOMAIN 655 717 COILED COIL (POTENTIAL).
 FT DOMAIN 717 799 COILED COIL (POTENTIAL).
 FT SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 82 91 BY SIMILARITY.
 FT DISULFID 84 98 BY SIMILARITY.
 FT DISULFID 101 110 BY SIMILARITY.
 FT DISULFID 113 129 BY SIMILARITY.
 FT DISULFID 132 146 BY SIMILARITY.
 FT DISULFID 134 155 BY SIMILARITY.
 FT DISULFID 157 186 BY SIMILARITY.
 FT DISULFID 169 184 BY SIMILARITY.
 FT DISULFID 187 202 BY SIMILARITY.
 FT DISULFID 189 209 BY SIMILARITY.
 FT DISULFID 212 221 BY SIMILARITY.
 FT DISULFID 224 238 BY SIMILARITY.
 FT DISULFID 266 266 INTERCHAIN (PROBABLE).
 FT DISULFID 269 269 INTERCHAIN (PROBABLE).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 524 524 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1086 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1281 1281 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1359 1359 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1411 1411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 143 143 A -> P (IN REF. 1).
 FT CONFLICT 178 178 L -> F (IN REF. 1).
 FT CONFLICT 265 265 G -> GMDPTIS (IN REF. 4).
 FT CONFLICT 276 276 D -> A (IN REF. 4).
 FT CONFLICT 491 491 Y -> H (IN REF. 2 AND 3).
 FT CONFLICT 1057 1057 T -> P (IN REF. 1).
 FT CONFLICT 1110 1112 SGR -> GGP (IN REF. 4).
 SQ SEQUENCE 1816 AA; 201908 MW; 04E9AF379A0F4A4D CRC64;
 Query Match 5.8%; Score 93; DB 1; Length 1816;
 Best Local Similarity 23.8%; Pred. No. 4.2;

Matches 74; Conservative 32; Mismatches 143; Indels 62; Gaps 18;

QY 2 VICQLKGAQMLCIDNCG-----ARELKALH---LLPOY--DDQSFPOSELPKPMPTTLV 51
 Db 80 VPCDCNGSNE-CLDGGYCVHCQRWTTGEHCEKCLDGYIGDSIRGAPOFCQPCP----- 133
 QY 52 GRLLPVPKLNLTQV--DNGALPSAVN--GAAPFS-----GPAQLQGPPKITLSGYCDCFSS 104
 Db 134 ---CPLPLHANAESYRKNAGVACICNENYAGPNCERCAPGYGPNPLI----- 180
 QY 105 GDFNCSCSN-----NLRIE--LERPKATKACLDNRPEAFQ-PKMGKGRGAAL-RHSK 155
 Db 181 GSTCKKDCSGNSDPLIFEDCDVTGOCRNCL-RNTTGFKCERCAPGYGDAIAKNCA 239
 QY 156 GNCCKRSGCLKNYCEYKINCSSTC-KCI-----ACKNYEESPERKMLMST--P 203
 Db 240 VMCGGGPGCDVTGCEGEGFPPTCDCKVWDLTDDLRLAALSIEEGSGVLSVSSGAA 299
 QY 204 HMEPGDFESSHYLSPAKFSGPPKLRKROAFSCISWEVVEATCACLQAQGEAEQEHCS 263
 Db 300 AHRHVEINATIVLLTKLSE---RENOYALRKIQINNAENTMKSLLSDVEELVEKENQ 355
 QY 264 PSIAEQMILEE 274
 Db 356 ASRGOLVQKE 366

RESULT 11

ITBL_XENLA STANDARD; PRT; 798 AA.

AC P12606;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin beta-1 precursor.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88186829; PubMed=2833505;
 RA Desimone D.W., Hynes R.O.;
 RT "Xenopus laevis Integrins. Structural conservation and evolutionary
 divergence of integrin beta subunits.";
 RL J. Biol. Chem. 263:5333-5340(1988).
 CC -!- FUNCTION: BETA INTEGRINS ASSOCIATE WITH ALPHA SUBUNITS TO FORM
 RECEPTOR COMPLEXES THAT RECOGNIZE THE SEQUENCE R-G-D IN A WIDE
 ARRAY OF LIGANDS (BY SIMILARITY).
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND BETA SUBUNIT.
 CC -!- SUBCELLULAR LOCATION: TYPE I membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY. THE
 INTEGRINS BETA-1 AND BETA-1* OF XENOPUS ONLY DIFFER BY 19 AMINO
 ACIDS.
 CC -!- SIMILARITY: CONTAINS 1 WFVA-LIKE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M20140; AAA49889.1; -
 CC PIR; A28193; A28193.
 CC HSSP; P05106; 1JW2.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR002369; Integrin_B.
 CC InterPro; IPR001169; Integrin_beta_C.
 CC InterPro; IPR003659; Plexin-like.
 CC InterPro; IPR002035; WFVA.

DR Pfam: PF00362; Integrin_B: 1.
 DR PRINTS; PR01186; INTEGRINB.
 DR PRODOM; PD001811; Integrin_B: 1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA: 3.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KW Repeat; Signal; Phosphorylation.
 FT SIGNAL 1 21
 FT CHAIN 22 798 INTEGRIN BETA-1.
 FT DOMAIN 22 727 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 728 751 POTENTIAL.
 FT DOMAIN 752 798 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 139 377 WFVA-LIKE.
 FT DOMAIN 466 635 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 466 515 I.
 FT REPEAT 516 559 II.
 FT REPEAT 560 598 III.
 FT REPEAT 599 635 IV.
 FT DISULFID 28 464 BY SIMILARITY.
 FT DISULFID 36 46 BY SIMILARITY.
 FT DISULFID 39 76 BY SIMILARITY.
 FT DISULFID 49 65 BY SIMILARITY.
 FT DISULFID 206 212 BY SIMILARITY.
 FT DISULFID 260 300 BY SIMILARITY.
 FT DISULFID 400 414 BY SIMILARITY.
 FT DISULFID 434 691 BY SIMILARITY.
 FT DISULFID 462 466 BY SIMILARITY.
 FT DISULFID 477 489 BY SIMILARITY.
 FT DISULFID 486 525 BY SIMILARITY.
 FT DISULFID 491 500 BY SIMILARITY.
 FT DISULFID 502 516 BY SIMILARITY.
 FT DISULFID 531 536 BY SIMILARITY.
 FT DISULFID 533 568 BY SIMILARITY.
 FT DISULFID 538 553 BY SIMILARITY.
 FT DISULFID 555 560 BY SIMILARITY.
 FT DISULFID 574 579 BY SIMILARITY.
 FT DISULFID 576 607 BY SIMILARITY.
 FT DISULFID 581 590 BY SIMILARITY.
 FT DISULFID 592 599 BY SIMILARITY.
 FT DISULFID 613 618 BY SIMILARITY.
 FT DISULFID 615 661 BY SIMILARITY.
 FT DISULFID 620 630 BY SIMILARITY.
 FT DISULFID 633 636 BY SIMILARITY.
 FT DISULFID 640 649 BY SIMILARITY.
 FT DISULFID 646 723 BY SIMILARITY.
 FT DISULFID 665 699 BY SIMILARITY.
 FT MOD_RES 783 783 PHOSPHORYLATION (BY SIMILARITY).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 798 AA; 88167 MW; 093522509B298531 CRC64;

Query Match

Best Local Similarity 5.8%; Score 92; DB 1; Length 798;

Matches 47; Conservative 14; Mismatches 71; Indels 82; Gaps 12;

QY 95 LSGYCD-----CFSSGD-FCNSCSNLRHELEFRKAKACLDNRNPEAFQPKMGKGR 146
 Db 512 MDAYCRRNSSEICSNNGDCICQCVCCKRDNPNVSG-KYCEDN---FNCDRNGLI 567
 QY 147 GAAKLRHSGK-CNCKRSGCLKNY---CECYEAKIMCSS-----ICKCIACKNVE 191

Db 568 CG-----GKGICKRVCEFPNYSGSACDSEDSTCMKNGQICNGRICDCGRCK--- 619

Qy 192 ESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNQAFSCISWEVVEATC-ACL 250

Db 620 -----CITDPKFGQP-----TCELQ 634

Qy 251 LAQGEAEQEHCSPLAQFQMLTEFGRCLSLILH 284

Db 635 TCVGVCAEHKECVOCRAFOK-GEKQDVCMEQCMH 667

RESULT 12

CD93 MOUSE

ID CD93_MOUSE STANDARD; PRT; 644 AA.

AC O89103;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE Complement component C1q receptor precursor (Complement component 1, q subcomponent, receptor 1) (C1qR) (C1qR(p)) (C1q/MBL/SPA receptor)

DE (CD93 antigen) (Cell surface antigen AA4) (Lymphocyte antigen 68).

GN C1Q1 OR CD93 OR C1QRP OR LY68 OR AA4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN RN SEQUENCE FROM N.A.

RP STRAIN=129/SV;

RC PubMed=11074255;

RX Kim T.S., Park M., Nepomuceno R.R., Palmirani G., Winokur S.,

RA Colman C.A., Bengtsson U., Tenner A.J.;

RA "Characterization of the murine homolog of C1qR(p): identical cellular

RT expression pattern, chromosomal location and functional activity of

RT the human and murine C1qR(p).";

RL Mol. Immunol. 37:377-389(2000).

RN [2]

RN SEQUENCE FROM N.A.

RP TISSUE=Leukemia;

RC PubMed=10403644;

RX Petrenko O., Beavis A., Klaine M., Kittappa R., Godin I.,

RA Lenischka I.R.;

RA "The molecular characterization of the fetal stem cell marker AA4.";

RT Immunity 10:691-700(1999).

RN [3]

RN SEQUENCE FROM N.A.

RP STRAIN=129/SV; TISSUE=Spleen, and Endothelial cells;

RC MEDLINE=99359842; PubMed=10430665;

RX Norsworthy P.J., Taylor P.R., Walport M.J., Botto M.;

RA "Cloning of the mouse homolog of the 126-kDa human C1q/MBL/SP-A

RT receptor, C1qRp.";

RL Mamm. Genome 10:789-793(1999).

CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for

CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant

CC protein A (SPA). May mediate the enhancement of phagocytosis in

CC monocytes and macrophages upon interaction with soluble defense

CC collagens. May play a role in intercellular adhesion. Marker for

CC early multipotent hematopoietic precursor cells. May play a role

CC in cell-cell interactions during hematopoietic and vascular

CC development.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- TISSUE SPECIFICITY: Expressed in lung, heart and bone marrow.

CC Expressed at lower level in ovary, whole embryo and fetal liver.

CC Not detected in brain, adult liver or thymus. Highly expressed in

CC peritoneal cavity and bone marrow macrophages. Not detected in

CC epithelial cells.

CC -!- DEVELOPMENTAL STAGE: First detectable in day 9 embryos, in the

CC endocardium and vascular endothelium in the anterior part of the

CC embryo. Expression in endothelial cells, initially restricted to

CC aorta, omphalomesenteric and umbilical arteries, later extends to

CC subcardinal veins, intersomitic arteries and perimeural vessels.

CC On day 10, detectable in the entire embryo.

CC -!- PTM: N- and O-glycosylated (by similarity):

CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC EMBL; AF074856; AAC63274.1; -.

DR EMBL; AF081789; AAC62649.1; -.

DR EMBL; AF099939; AAD47906.1; -.

DR EMBL; AF099938; AAD47906.1; JOINED.

DR HSSP: P35555; 1EMN

DR MGD; MGI:106664; Ly68.

DR InterPro: IPR000152; Asx_hydroxyl.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR001881; EGF_Ca.

DR InterPro: IPR001304; Lectin_C.

DR Pfam: PF00008; EGF_5.

DR Pfam: PF00059; lectin_C; 1.

DR SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_CA; 3.

DR SMART; SM00001; EGF_like; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 3.

DR PROSITE; PS00615; C-TYPE_LECTIN_1; FALSE_NEG.

DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.

DR PROSITE; PS01186; EGF_2; 3.

DR PROSITE; PS01187; EGF_CA; 3.

KW Receptor; EGF-like domain; Signal; Transmembrane; Glycoprotein;

KW Repeat; Lectin.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 644 COMPLEMENT COMPONENT C1Q RECEPTOR.

FT DOMAIN 23 572 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 573 593 POTENTIAL.

FT DOMAIN 594 644 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 31 173 C-TYPE LECTIN.

FT DOMAIN 257 298 EGF-LIKE 1.

FT DOMAIN 299 341 EGF-LIKE 2.

FT DOMAIN 342 381 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 382 423 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 424 465 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).

FT DISULFID 261 272 BY SIMILARITY.

FT DISULFID 268 282 BY SIMILARITY.

FT DISULFID 284 297 BY SIMILARITY.

FT DISULFID 303 314 BY SIMILARITY.

FT DISULFID 308 325 BY SIMILARITY.

FT DISULFID 327 340 BY SIMILARITY.

FT DISULFID 346 355 BY SIMILARITY.

FT DISULFID 351 364 BY SIMILARITY.

FT DISULFID 366 380 BY SIMILARITY.

FT DISULFID 386 397 BY SIMILARITY.

FT DISULFID 393 406 BY SIMILARITY.

FT DISULFID 408 422 BY SIMILARITY.

FT DISULFID 428 440 BY SIMILARITY.

FT DISULFID 436 449 BY SIMILARITY.

FT DISULFID 451 464 BY SIMILARITY.

FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 322 322 N-LINKED (GLCNAC...) (POTENTIAL).

SQ SSEQUENCE 644 AA; 69354 MW; EB4351648BF8635A CRC64;

Query Match 5.7%; Score 91.5; DB 1; Length 644;

Best Local Similarity 24.8%; Pred. No. 1.8;

Matches 37; Conservative 12; Mismatches 51; Indels 49; Gaps 8;

Qy 83 SGPALGPPKTIIS---GYC--DCFSSGDFCNSSCNLRHELFERKAIKACLDNRPEAF 137

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 250 SGPVSPFGGSGFNNNGCQDFEGGDSFRGCG---RPGFRLDLVTCASRNPCSS 305

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 138 QPKMGKRLGAALKRHSKGC-----NCKRSGCLKN-----YCE 170

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||


```

Db 306 NPCTGGMCHSVPLSENVTCTCPGQYQLDSQVHCVIDECQDSPCAQCVNTLGSFHC 365
QY 171 CY-----EAKIMSSICKCIACKN 189
Db 366 CWVGQPSGPKKEA---CEDVDECAANS 391

RESULT 13
OC90_HUMAN
ID OC90_HUMAN STANDARD; PRT; 493 AA.
AC Q02509;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Otc90in 90 precursor (Otc90) (Phospholipase A2 homolog).
GN OC90 OR PLA2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93181166; PubMed=8382789;
RA Feuchter-Murthy A.E., Freeman J.D., Mager D.L.;
RT "Splicing of a human endogenous retrovirus to a novel phospholipase
RT A2 related gene."
RL Nucleic Acids Res. 21:135-143(1993).
CC -!- FUNCTION: IT IS UNLIKELY THAT THIS PROTEIN HAS PHOSPHOLIPASE A2
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. CONTAINS 3
CC PA2 TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z14310; CAA78662.1; ALT_INIT.
DR HSSP; P00593; 4BP2.
DR Genew; HGNC:8100; OC90.
DR MIM; 601658.
DR InterPro; IPR001211; PhospholipaseA2.
DR Pfam; PF00068; phoslip; 2.
DR PRINTS; PR00389; PHPLIPASEA2.
DR ProDom; PD000303; PhospholipaseA2; 2.
DR SMART; SM00085; PA2c; 2.
DR PROSITE; PS00118; PA2_HIS; 2.
DR PROSITE; PS00119; PA2_ASP; 1.
KW Signal; Glycoprotein; Repeat.
FT SIGNAL 1 17
FT CHAIN 18 493
FT DOMAIN 76 190
FT DOMAIN 321 377
FT DOMAIN 389 441
FT DISULFID 85 145
FT DISULFID 99 190
FT DISULFID 101 117
FT DISULFID 116 172
FT DISULFID 123 165
FT DISULFID 132 158
FT DISULFID 152 163
FT CARBOHYD 38 38
FT CARBOHYD 179 179
FT CARBOHYD 423 423
SQ SEQUENCE 493 AA; 53338 MW; 59F9AF7C1364A5B7 CRC64;

Query Match 5.7%; Score 91; DB 1; Length 493;
Best Local Similarity 18.3%; Pred. No. 1.5;
Matches 59; Conservative 50; Mismatches 89; Indels 124; Gaps 14;

```

```

QY 2 VICOLKGGQA-MLCIDNCGARELKA-----LHLLPQYDDOSSPQSELPKPMITLVCR 53
Db 150 IICESKDNCEHLLCTCDKAAIECLARSSUNSLNLLDTSCFLAQPTPETTKEDLTLLR 209
QY 54 LILPV-PAKLNLI-----TQVD-----NGALPSAVNGAA-----80
Db 210 VVPVEPTDTSLTALSSEVAATEADRLITLSKKAGHDQEGVCAARATSPPGSAEIVATR 269
QY 81 -----FPSG-----PALQPPKPTLGYDCDFSSGDFCNSCNSNNRHELERPK 124
Db 270 VTAKIIVTLVPAGIKSLGLAVSSVENDPEETTEKACDRFT---FLHLSGDNQVMPQLGE 326
QY 125 AIKACILDRNPEAFQPKMGKRLGAAKLHSGKCNKRSQ-----CLKNYCECYE 173
Db 327 MLFCLTSTRCPPEFE-----SYGICYGQEGRPRDDLRCCLSLHHC-CUE 370
QY 174 -----AKTMS-----SICKCIACKNYEESPERKM 198
Db 371 QVRRLGCLLERLPWSPVVCVDHTPKCGGQSLCKELLCACDQTAACMTSASFNSQLKSPS 430
QY 199 LMSTPHYMEPGDFESSHYLSPA 220
Db 431 RLGCFC--QPAACEDSLHPVA 450

RESULT 14
EZHL_HUMAN
ID EZHL_HUMAN STANDARD; PRT; 747 AA.
AC Q92800; O43287; Q14459;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Enhancer of zeste homolog 1 (ENX-2).
GN EZHL OR KIAA0388.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97079676; PubMed=8921387;
RA Abel K.J., Brody L.C., Valdes J.M., Erdos M.R., McKinley D.R.,
RA Castilla L.H., Merajver S.D., Couch F.J., Friedman L.S.,
RA Ostermeyer E.A., Lynch E.D., King M.-C., Welcsh P.B.,
RA Osborne-Lawrence S., Spillman M., Bowcock A.M., Collins F.S.,
RA Weber B.L.;
RT "Characterization of EZHL, a human homolog of Drosophila Enhancer of
RT zeste near BRCAL."
RL Genomics 37:161-171(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98146265; PubMed=9473645;
RA Ogawa M., Hiraoka Y., Taniguchi K., Aiso S.;
RT "Cloning and expression of a human/mouse Polycomb group gene,
RT ENX-2/Enx-2."
RL Biochim. Biophys. Acta 1395:151-158(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [5]

```


Db 657 TPGICMNGHCINNEGSRFCDCPPGLAVGVDGRVCVDTNMRSTCYGEIKKGVCVRPFPGAV 716
Qy 249 ----CLLAQGEAEQEHCSPLAEQMIILEEFGRLSQI 282
Db 717 TKSECCCANPDYGRGEPQCPKANS-AEFHGLCSSGI 753

Search completed: April 21, 2003, 11:48:56
Job time : 12.9394 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 21, 2003, 11:30:23 ; Search time 32.7778 Seconds
(without alignments)
1199.256 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRLCSQILHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1599	100.0	295	21	Mouse testis speci
2	1251	78.2	299	21	Human testis speci
3	435	27.2	950	22	Drosophila melanog
4	396	24.8	438	20	Caenorhabditis ele
5	362	22.6	280	22	Novel human diagn
6	322	20.1	147	22	Human testicular a
7	322	20.1	147	22	Human reproductive
8	248	15.5	53	22	Human colon cancer
9	186	11.6	243	22	Drosophila melanog
10	165	10.3	251	22	Human polypeptide,

11	110.5	6.9	732	18	AAW26642	Human RECK cancer-
12	99	6.2	2931	22	ABB68229	Drosophila melanog
13	97	6.1	602	21	AAV95660	Human Zntn2 protei
14	97	6.1	1987	22	ABB61083	Drosophila melanog
15	96.5	6.0	359	21	AAG09497	Arabidopsis thalia
16	96.5	6.0	443	21	AAG09496	Arabidopsis thalia
17	96.5	6.0	472	21	AAG09495	Arabidopsis thalia
18	96.5	6.0	1788	22	ABB62995	Drosophila melanog
19	96.5	6.0	4547	22	ABB59051	Drosophila melanog
20	96	6.0	773	22	ABB59751	Drosophila melanog
21	96	6.0	3635	23	ABB81589	Mouse laminin alph
22	96	6.0	3635	23	AAW50357	Mouse laminin-15 a
23	95.5	6.0	1035	22	ABB66062	Drosophila melanog
24	94.5	5.9	759	23	AAU97037	Human LP protein L
25	94	5.9	1792	21	AAW48443	Human laminin 8 po
26	94	5.9	1800	21	AAW48445	Human laminin 8 po
27	94	5.9	1816	21	AAW48442	Human laminin 8 po
28	94	5.9	1824	21	AAW48444	Human laminin 8 po
29	93.5	5.8	652	19	AAW49879	Amino acid sequenc
30	93.5	5.8	652	21	AAW32345	Human cell surface
31	93.5	5.8	1981	22	ABB61657	Drosophila melanog
32	92.5	5.8	921	21	AAV32297	Corn polycomb prot
33	92	5.8	3680	22	ABB70878	Drosophila melanog
34	91.5	5.7	637	21	AAV32346	Mouse cell surface
35	91.5	5.7	644	21	AAV79186	Haematopoietic ste
36	91.5	5.7	644	21	AAV79193	Haematopoietic ste
37	91	5.7	18	21	AAV68465	Mouse testis speci
38	90	5.6	533	23	ABB53284	Human polypeptide
39	89.5	5.6	347	22	AAW87443	Human gene 35 enco
40	89.5	5.6	361	22	AAW99918	Human polypeptide
41	89.5	5.6	439	22	AAW95609	Human protein sequ
42	89	5.6	1725	21	AAW19800	Mouse laminin 2 ma
43	89	5.6	1725	21	AAW48451	Mouse laminin 8 po
44	89	5.6	1725	23	ABB81593	Mouse laminin 10 s
45	89	5.6	1764	10	AAW91672	Primary amino acid

ALIGNMENTS

RESULT 1

AAV68463
ID AAV68463 standard; Protein; 295 AA.

XX AAV68463;

XX 25-APR-2000 (first entry)

XX Mouse testis specific factor tesmin SEQ ID NO:4.

XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
XX differentiation regulatory factor; male germ cell regulatory actor;
XX germ cell differentiation; sterility.

XX Mus musculus.

XX WO200004147-A1.

XX 27-JAN-2000.

XX 16-JUL-1999; 99WO-JP03859.

XX 17-JUL-1998; 98JP-0219856.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX (AGEN-) AGENCY OF IND SCI & TECHNOLOGY.

XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX WPI; 2000-147785/13.

XX N-PSDB; AAZ88155, AAZ88157.

XX New male germ cell regulatory factor tesmin expressed in spermatocytes

PT useful for investigation of germ cell differentiation and sterility
XX
PS Claim 1; Page 47-49; 63pp; Japanese.
XX
CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin.
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 295 AA;
Query Match 100.0%; Score 1599; DB 21; Length 295;
Best Local Similarity 100.0%; Pred. No. 3e-144;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTLVGRLLPVPK 60
DB 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTLVGRLLPVPK 60
QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKITLSGYDCDFSSGDFCNSCNNLRHEL 120
DB 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKITLSGYDCDFSSGDFCNSCNNLRHEL 120
QY 121 ERKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRKRSGLKKNYCEYEAQIMCSS 180
DB 121 ERKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRKRSGLKKNYCEYEAQIMCSS 180
QY 181 ICICIAKNYEEPERKMLMSTHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 240
DB 181 ICICIAKNYEEPERKMLMSTHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 240
QY 241 EVVEATCACLAAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 295
DB 241 EVVEATCACLAAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 295
RESULT 2
AAY68464
ID AAY68464 standard; Protein; 299 AA.
XX
AC AAY68464;
XX
XX
XX 25-APR-2000 (first entry)
DT Human testis specific factor tesmin SEQ ID NO:5.
DE
DE
XX
XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
XX Homo sapiens.
XX
XX WO200004147-A1.
PN
XX
XX 27-JAN-2000.
PD
XX
XX 16-JUL-1999; 99WO-JP03859.
PF
XX
XX 17-JUL-1998; 98JP-0219856.
PR
XX
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PA
XX
XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
PI
XX
XX WPI; 2000-147785/13.
DR
XX
XX N-PSDB; AA288156.
DR
XX
XX New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility
XX
XX Claim 1; Page 50-52; 63pp; Japanese.
PS

CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin.
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 299 AA;
Query Match 78.2%; Score 1251; DB 21; Length 299;
Best Local Similarity 76.3%; Pred. No. 5.1e-111;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;
QY 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTLVGRLLPVPK 60
DB 1 MVICQLKGGAGMCLIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTLVGRLLPVPK 60
QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKITLSGYDCDFSSGDFCNSCNNLRHEL 116
DB 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKITLSGYDCDFSSGDFCNSCNNLRHEL 120
QY 117 RHELEKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRKRSGLKKNYCEYEAQI 176
DB 121 HHDIERKAIKACLDNRNPEAFQPKMGKRGRLGAALKRHSKGCNCRKRSGLKKNYCEYEAQI 180
QY 177 MCSSICKICIAKNYEEPERKMLMSTHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFS 236
DB 181 MCSSICKICIAKNYEEPERKMLMSTHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFS 240
QY 237 CISWEVVEATCACLAAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 295
DB 241 CISWEVVEATCACLAAQGEAEQHCPSLAEQMILEEFGRCLSQILHIEFKSKGLKIE 299
RESULT 3
ABB62035
ID ABB62035 standard; Protein; 950 AA.
XX
AC ABB62035;
XX
XX 26-MAR-2002 (first entry)
DT
DE
DE Drosophila melanogaster polypeptide SEQ ID NO 12897.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX N-PSDB; ABL06138.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC Insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 950 AA;
 Query Match 27.2%; Score 435; DB 22; Length 950;
 Best Local Similarity 38.4%; Pred. No. 1.7e-32;
 Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;
 QY 31 QYDQSQFFQSELPKPTTLVGRLLPVPKLNLTQVDNGALPSAVNGAAPSPGALQGP 90
 DB 681 QLPTEQSTPIKVEKLTLPQGVKAVPAK--PLFEVLKPPATAAAGAVDPLGGMTSRR 738
 QY 91 PKITLSG-----YDCDFSSGDFCNSCSC-----NNLRHELERFKAIKACLDNRNPEAFQPK 140
 DB 739 KHCNCKSQCLKLYCDCFANGFQCQCTCKDCFNLDYEVERAIRSCLDNRNPSAFKPK 798
 QY 141 MGKRLGAAGAKLRHSGKNCNRSGCLKNYCEYEAKEIMCSTCKCIACKNYEESPERKMLM 200
 DB 799 ITAPNSGDMRL-HNKGNCNRSGCLKNYCEYEAKEIPCSISICKCVGCRNMDRPD----- 852
 QY 201 STPHMEPGDPFESSHYLSPAKFSGPPKLRNRO-----AFSCISWEVVEATCCLLAQGE 255
 DB 853 -----VMDSLDGLMGVGGQKDK-ARNKQLNENRANIYFTDDVIEATIMCISRIV 903
 QY 256 EAEQHCSPSLAEQMIIEFGRCILSQIL 283
 DB 904 MHEQKNVAVEDMEREVMEEMGESLTQII 931
 RESULT 4
 AAW83392
 ID AAW83392 standard; Protein; 438 AA.
 XX
 AC AAW83392;
 DT 29-MAR-1999 (first entry)
 XX
 DE Caenorhabditis elegans synMuv protein LIN-54.
 XX
 KW LIN-54; synthetic multivulvar; SynMuv; signal transduction;
 KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
 KW cell proliferation; gene therapy.
 XX
 OS Caenorhabditis elegans.
 XX
 PN WO9854299-A1.
 XX
 PD 03-DEC-1998.
 XX
 PF 28-MAY-1998; 98WO-US11043.
 XX
 PR 28-MAY-1997; 97US-0047996.
 XX
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 PI Ceol C, Horvitz HR, Lu X;
 XX
 DR WPI; 1999-045362/04.
 DR N-PSDB; AAV72865.
 XX
 PT Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
 PT useful for treating diseases associated with altered levels of cell
 PT proliferation, e.g. carcinomas
 XX
 PS Claim 7; Fig 13; 70pp; English.
 XX
 CC This is the amino acid sequence of LIN-54, a novel protein of

CC Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
 CC synthetic multivulvar (synMuv) gene involved in cell fate and cell
 CC proliferation, and is part of a pathway that may be used as a
 CC genetic and biochemical model system for tumour suppression and
 CC cancer in mammals. SynMuv pathway genes and proteins may be used
 CC to identify genes which are part of the mammalian pathway and
 CC to identify genes, proteins and therapeutic compounds which
 CC modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
 CC C. elegans SynMuv polypeptides selected from LIN-37, -35, -55, -52,
 CC -53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
 CC (1) nucleic acids (see AAV72866-67) encoding the mouse and human
 CC homologues of C. elegans LIN-54; (2) vectors containing the nucleic
 CC acids; (3) transgenic cells; (4) a pure mammalian SynMuv
 CC polypeptide, and (5) an antibody which binds to a SynMuv family
 CC protein. The SynMuv nucleic acids and polypeptides can be used to
 CC diagnose and treat, especially by gene therapy, conditions
 CC involving altered levels of cell proliferation, e.g. SynMuv-
 CC associated carcinomas.
 XX
 SQ Sequence 438 AA;
 Query Match 24.8%; Score 396; DB 20; Length 438;
 Best Local Similarity 38.9%; Pred. No. 3.4e-29;
 Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;
 QY 98 YDCDFSSGDFCNSCSC-----NNLRHELERFKAIKACLDNRNPEAFQPKMGRIGAAKLR- 152
 DB 191 YDCDFANGFRCNCKDCHNNIETDSQSKAIRQSLERNPNAEPKPIGIARGGTIDIER 250
 QY 153 -HSKGCNCKRSGCLKNYCEYEAKEIMCSTCKCIACKN-----YEE-----PERKML 199
 DB 251 LHQSGCHCKRSGCLKNYCEYEAKEVPCTDRCKCKGCQNTETRYMTRKNSGGAVSNTNAL 310
 QY 200 MS-----TPHYMEPGDPFESSHY-----LSPAKFSGPPKLRNROAFSCISN-----EVV 243
 DB 311 MSFTNASSTATPD-SGPGSVVTVDEHGDYEDMLLSHKPKVEMDPRPF---PWYTMDEV 366
 QY 244 EATCACLLAQGEA-----BOEHCSPLSABOMILEEFGRCILSQIL 283
 DB 367 EATMCMVAQAEALNYEKVQTEDEKLNMEKLVLEFGRCLEQMI 412
 RESULT 5
 ABG17958
 ID ABG17958 standard; Protein; 280 AA.
 XX
 AC ABG17958;
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #17949.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS82145.
 XX

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.
XX
XX

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

N-PSDB; AAL01300.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a protein of the invention.

Sequence 147 AA;

Query Match 20.1%; Score 322; DB 22; Length 147;
Best Local Similarity 73.8%; Pred. No. 9.7e-23;
Matches 62; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDQSSFPQSELKPKMTTLVGRLLPVPK 60

Db 60 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDQSSFPQSELKPKMTTLVGRLLPVPK 119

QY 61 LNLITVDNGALPSAVNGAAPPSG 84

Db 120 LNLITQLEALPSVNGAAPPSG 143

RESULT 8

AAG76158

ID AAG76158 standard; Protein; 53 AA.

XX AAG76158;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6922.

XX

XX 07-JUL-2000; 2000EP-0114089.
 XX
 XX 08-JUL-1999; 95JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94268.
 XX
 XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PT
 XX
 PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
 CC
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 251 AA;
 Query Match 10.3%; Score 165; DB 22; Length 251;
 Best Local Similarity 75.6%; Pred. No. 1.9e-07;
 Matches 31; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MVICQLKGGAOMLCIDNCGARELKALHLLPOYDDQSSFPQS 41
 DB 210 MVICQLKGGAOMLCIDNCGARELKALHLLPOYDDQSSFPQS 250
 RESULT 11
 ID AAW26642
 XX
 XX AAW26642 standard; Protein; 732 AA.
 AC AAW26642;
 XX
 DT 11-FEB-1998 (first entry)
 XX
 DE Human RECK cancer-inhibiting protein.
 XX
 KW RECK; reversion-inducing cysteine rich protein with Kazal motif;
 KW human; cancer; gene therapy; diagnosis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 396..415
 FT /label= Kazal_domain
 XX
 PN WO9724439-A1.
 XX
 PD 10-JUL-1997.
 XX
 PF 24-DEC-1996; 96WO-US20812.
 XX
 PR 27-DEC-1995; 95JP-0340469.
 XX
 PA (AMGE-) AMGEN INC.
 PA (KITA/) KITAYAMA H.

PA (NODA/) NODA M.
 PA (TAKA/) TAKAHASHI C.
 PA (SANY) SANKYO CO LTD.
 XX
 PI Kitayama H, Noda M, Takahashi C;
 XX
 DR WPI; 1997-363675/33.
 DR N-PSDB; AAT90508.
 XX
 XX RECK gene and corresponding protein sequences - enables reversion of
 PT cancer cells
 XX
 PS Claim 7; Page 42-44; 53pp; English.
 CC
 CC This protein sequence comprises the human RECK protein (reversion-
 CC inducing cysteine rich protein with Kazal motif), which is capable
 CC of transforming malignant cancer cells with an activated ras gene
 CC into normal cells (reversion activity). Its amino acid sequence
 CC was deduced from a cDNA clone (see AAT90508) obtained from human
 CC fibroblast MRC-5 (ATCC CCL 171) cells. A claimed method for
 CC treatment of cancer comprises contacting the cancer cells with a
 CC RECK polypeptide. RECK may also be expressed using gene therapy
 CC methods for in vivo treatment of cancer.
 XX
 SQ Sequence 732 AA;
 Query Match 6.9%; Score 110.5; DB 18; Length 732;
 Best Local Similarity 17.9%; Pred. No. 0.12;
 Matches 63; Conservative 41; Mismatches 110; Indels 137; Gaps 17;
 QY 4 COLKGGAMLCIDNCGARELKALHLLPOYDDQSSFPQSSELKPKMTLVG-----RLLPV 57
 DB 183 CHSKSRGSIICKSDC-----VEILKKCGDQNKPPEDHTAESICELLSPTDDLKNCIPL 235
 QY 58 -----PAKL-NLITQVDNGLPS-----AVNGAATPSG-PALQGPKITLSGYCDGF 102
 DB 236 DTYLRPSTLGNIVEEVEHPNPNPCANELCEVNRKKGCPGSDPCL---PYFCVQG-CKLG 291
 QY 103 SSGDFCNCSNNLRHELERFKAICACLDNRNPEAFQPKMKGRLGAALRHSKGCNCKRS 162
 DB 292 EASDF-----IVROGTLQVPSASAGEVCYKI-----CSCQGS 324
 QY 163 GCLK-----NYCEYEAKIMCSS----- 180
 DB 325 GLLNCMEMHCHIDLQKSCIVGGRKRSHTSFSIDCNVCSFAGNLVCTRLCLSEHSSD 384
 QY 181 -----ICKCI-----AC---KNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFS 223
 DB 385 DRRFTTGLPCNCAQDFVPVCGQNGRTYPSACIARCVGLQDHQHFEGSCMSK---DPC--- 438
 QY 224 GPPKLRKNRQA-----FSCISWEVVEATCACLIAQGEARQEH 261
 DB 439 NPNCKQKQRCIPKQVCLTTFDKFGCSQYECVPRQLACDQVQDPVCDTDH 489
 RESULT 12
 ID ABB68229
 ID ABB68229 standard; Protein; 2931 AA.
 XX
 AC ABB68229;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 31479.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.
XX
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX
PA (PEKE) PE CORP NY.
XX
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX
DR WPI; 2001-658860/75.
DR N-PSDB; ABL12332.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
XX
PS Disclosure; SEQ ID NO 31479; 21pp + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 2931 AA;

Query Match 6.2%; Score 99; DB 22; Length 2931;
Best Local Similarity 21.3%; Pred. NO. 8.6;
Matches 50; Conservative 26; Mismatches 81; Indels 78; Gaps 11;

QY 100 DCFSSGDFCNCS-----CNLRLHLERFKAKACLDNRNPAFQPK 140
DB 1677 DCASSGGTLTKCGYDNCVTFVSDSGSVTQRCSE-----SVFEESDYC-DENP-AYCPR 1729

QY 141 MCKGRGLGAALKRSGCNCNKRSGCLKNYCEYKIMCISIC-----KCI 185
DB 1730 C-----NSNGCNTADS--QDKYVEICVDSSVDSNCVSDPTQITKTROCHERCI 1776
QY 186 AC-----KNYESPERKMLMSTPHYEPGDFESSHYLSPAKFGSPKLRKNRQAFSCISWE 241
DB 1777 SAFPLPNETEDPSYALIRNCYDLDLEKEDRD-----ACTAGSKRFCATCGTGCNSD 1829

QY 242 VYEATCACLQAQEEA-----EQEHCSPLAFQOMILEEFGRCLSOILH 284
DB 1830 LVASRHSCLVCGDECCQSQPQSSCSNYRHHDECYIOFDEERSITSLG-CLSELSH 1883

RESULT 13
AAV95660
ID AAV95660 standard; Protein; 602 AA.
XX
AC AAV95660;
XX
XX
DT 25-OCT-2000 (first entry)
XX
DE Human Zntn2 protein.
XX
XX Zntn2; epidermal growth factor-like domain; human;
KW cell differentiation; vulnery; diagnosis; therapy.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= Signal_peptide
FT /note= "the signal peptide alternatively

FT Protein 31..602 comprises amino acid residues 9..30"
FT Domain 31..507
FT /note= "extracellular domain"
FT Domain 508..533
FT /note= "transmembrane domain"
FT Domain 534602
FT /note= "intracellular domain"
FT Region 224..256
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 224..254 or 224..251"
FT Region 272..303
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 272..301 or 272..298"
FT Region 317..351
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 317..349 or 317..346"
FT Region 371..402
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 317..400 or 317..397"
FT Region 422..449
FT /note= "laminin-type epidermal growth factor-like
FT domain signature"
XX
XX WO200043512-A1.
PN 27-JUL-2000.
XX
XX
PF 20-JAN-2000; 2000WO-US01419.
XX
PR 25-JAN-1999; 99US-0237074.
XX
XX (ZYMO) ZYMOGENETICS INC.
XX
XX Holloway JL, Lofton-Day CE, Gilbert T;
XX WPI; 2000-491163/43.
DR N-PSDB; AAA50101, AAA50102.
XX
XX Isolated Zntn2 nucleic acids and polypeptides which act as epidermal
XX growth factors, useful for the treatment of e.g. kidney and liver
XX disorders, burns, and ulcers and for regulating smooth muscle cell
XX proliferation
XX
XX Claim 1; Page 89-91; 98pp; English.
XX
XX The present sequence is of that of human Zntn2, a new member of the
XX epidermal growth factor-like domain group of proteins. The
XX sequence was deduced from isolated Zntn2 cDNA (see AAA50101). In
XX addition to the presence sequence, polypeptides comprising amino
XX acids 31-507 and 31-602 of the sequence are claimed. An expression
XX vector, a recombinant host cell, a method of using the vector to
XX produce Zntn2 protein, an antibody or antibody fragment that
XX specifically binds to the polypeptide, and a method of detecting
XX the presence of Zntn2 gene products in a biological sample are also
XX all claimed. Zntn2 polypeptides may be used to regulate vascular
XX smooth muscle cell proliferation, to restore normal neurological
XX functioning after trauma, to treat ocular disorders, to treat
XX kidney and liver disorders, to promote hair and follicular
XX development, to stimulate growth and differentiation of various
XX epidermal and epithelial cells in vivo and in vitro, for the
XX treatment of burns, ulcers and corneal incisions, and to stimulate
XX wound healing.
SQ Sequence 602 AA;

Query Match 6.1%; Score 97; DB 21; Length 602;
Best Local Similarity 21.8%; Pred. No. 1.8;
Matches 56; Conservative 26; Mismatches 101; Indels 74; Gaps 16;

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155386.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 6.0%; Score 96.5; DB 21; Length 359;
 Best Local Similarity 32.4%; Pred. No. 1;
 Matches 34; Conservative 10; Mismatches 38; Indels 23; Gaps 7;

QY 131 DRNPEAFQPKMGKRLGAALRHSG---CNC-KRSGCLKNYCEYEAKIMCSSICKCIA 186
 : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
 Db 238 ESNSEFPQPTSETL--CKKEQSKAEVCCCTKKSICKTKCKCKANGSGCGDSCGCLA 295
 QY 187 --CKNYEES--PERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPK 227
 : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
 Db 296 SKCSNREESAKDPK-----MEPIDVK-----KPAGISHDDK 327

Search completed: April 21, 2003, 11:48:30
 Job time : 35.7778 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:48:39 ; Search time 13.4091 Seconds
(without alignments)
647.304 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGGAQMLCIDNCGA.....GRCLSQLHILIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	93.5	5.8	552	2	US-08-751-305-2	Sequence 2, Appli
2	87.5	5.5	610	1	US-08-365-470-3	Sequence 3, Appli
3	87.5	5.5	610	3	US-09-209-668-19	Sequence 19, Appl
4	87.5	5.5	610	4	US-09-009-490A-89	Sequence 89, Appl
5	86.5	5.4	418	4	US-08-795-430-13	Sequence 13, Appl
6	86.5	5.4	418	4	US-09-355-700-13	Sequence 13, Appl
7	86.5	5.4	610	6	5217870-2	Patent No. 5217870
8	86.5	5.4	830	1	US-08-110-158-4	Sequence 4, Appli
9	85.5	5.3	769	2	US-08-789-078-1	Sequence 1, Appli
10	85.5	5.3	769	2	US-08-752-633-1	Sequence 1, Appli
11	85.5	5.3	769	2	US-08-476-062A-45	Sequence 45, Appl
12	85.5	5.3	769	2	US-07-728-215-31	Sequence 31, Appl
13	85.5	5.3	769	4	US-08-938-085A-31	Sequence 31, Appl
14	85.5	5.3	769	5	PCT-US95-04886-1	Sequence 1, Appli
15	85.5	5.3	769	5	PCT-US96-01314-45	Sequence 45, Appl
16	85.5	5.3	217	4	US-09-602-543-5	Sequence 5, Appli
17	85.5	5.3	235	4	US-09-602-543-4	Sequence 4, Appli
18	85.5	5.3	676	3	US-08-630-172-10	Sequence 10, Appl
19	85.5	5.3	676	4	US-09-375-419-10	Sequence 10, Appl
20	84.5	5.3	696	3	US-08-899-437-23	Sequence 23, Appl
21	84.5	5.3	696	4	US-09-126-121-23	Sequence 23, Appl
22	84.5	5.3	720	3	US-08-899-437-6	Sequence 6, Appli
23	84.5	5.3	720	4	US-09-126-121-6	Sequence 6, Appli
24	83.5	5.2	350	2	US-08-999-811-4	Sequence 4, Appli
25	83.5	5.2	350	2	US-08-824-996-2	Sequence 2, Appli
26	83.5	5.2	350	3	US-09-042-105-4	Sequence 4, Appli
27	83.5	5.2	350	4	US-08-510-133A-33	Sequence 33, Appl

28	83.5	5.2	350	4	US-08-585-895-33	Sequence 33, Appl
29	83.5	5.2	419	2	US-08-999-811-2	Sequence 2, Appli
30	83.5	5.2	419	3	US-09-042-105-2	Sequence 2, Appli
31	83.5	5.2	419	3	US-09-042-105-18	Sequence 18, Appl
32	83.5	5.2	419	4	US-08-795-430-8	Sequence 8, Appli
33	83.5	5.2	419	4	US-08-510-133A-35	Sequence 35, Appl
34	83.5	5.2	419	4	US-09-355-700-8	Sequence 8, Appli
35	83.5	5.2	419	4	US-09-355-700-58	Sequence 58, Appl
36	83.5	5.2	419	4	US-08-601-132-33	Sequence 33, Appl
37	83.5	5.2	419	4	US-08-706-054A-3	Sequence 3, Appli
38	83.5	5.2	419	5	PCT-US96-09001-2	Sequence 2, Appli
39	81.5	5.1	1964	4	US-09-467-997-1	Sequence 1, Appli
40	81.5	5.1	415	4	US-08-795-430-11	Sequence 11, Appl
41	81.5	5.1	415	4	US-09-355-700-11	Sequence 11, Appl
42	81.5	5.1	909	4	US-09-013-895A-4	Sequence 4, Appli
43	81.5	5.1	909	4	US-09-448-868-4	Sequence 4, Appli
44	81.5	5.1	2476	2	US-08-276-967-2	Sequence 2, Appli
45	80.5	5.0	3224	2	US-08-705-660-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-08-751-305-2
; Sequence 2, Application US/08751305
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Tenner et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-305-2

Query Match	5.8%;	Score 93.5;	DB 2;	Length 652;
Best Local Similarity	27.0%;	Pred. No. 0.66;		
Matches	30;	Conservative	11;	Mismatches 51;
				Indels 19;
				Gaps 4;
QY	96	SGVC--DCFSGGDFCNSCNNLRHLEFKAKACLDNRPEAFQPMKGRGLGAALRH	153	
Db	268	NGCHQDCEGGGDSFLCGC---RPGFRLDDLVTCASRNPCSSSPGCG-----	313	
QY	154	SKGCNCRSGCKLKNY-CECYEARIMC-SSICKICIAKNYEESEPKMLMSTP	203	
Db	314	--GATCVLPHGKNYTCRCPOGQVQLDSSQLDCVDVDECQDSPCAQECVNTF	362	

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/009,490A
;; FILING DATE: January 20, 1998
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 440,740
;; FILING DATE: May 12, 1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 063,167
;; FILING DATE: May 17, 1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 969,151
;; FILING DATE: February 10, 1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 007,997
;; FILING DATE: January 20, 1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 939,855
;; FILING DATE: September 2, 1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 567,286
;; FILING DATE: August 14, 1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jane Massey Licata
;; REGISTRATION NUMBER: 32,257
;; REFERENCE/DOCKET NUMBER: ISPH-0268
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (609) 810-1515
;; TELEFAX: (609) 810-1454
;; INFORMATION FOR SEQ ID NO: 89:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 610
;; TYPE: Amino Acid
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
;; ANTI-SENSE: no
US-09-009-490A-89

Query Match 5.5%; Score 87.5; DB 4; Length 610;

Best Local Similarity 21.6%; Pred. No. 2.4; Mismatches 119; Indels 117; Gaps 20;

Matches 74; Conservative 33; Mismatches 119; Indels 117; Gaps 20;

QY 15 IDNCGARE-----LKAHLHPQYDDQS-----SFQSELPKPMWTL-----VGR-LLPVPA 59
DB 177 IVNCTALESPEHSLVCSHPLGNSFNSSCSISDCRGYLPSSNETMQCMSSGSEWSAPIPA 236
QY 60 ----KLLITQVDNGALPSAVNGAAPP-----SGPALQGGPKITLGYDCDFSSG 105
DB 237 CNVVECDVATNPANGFVECFQNPFGSPFWNTCTFDCGEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLHELEFRKAIKACLDNRNPEAFQPKMGKGRLG---AAKLHRSKG 156
DB 291 NMDNEKPTCKAVTCRAVR-----OPQNGSVRCSSHPAGEFTFKSS 330
QY 157 CN--CKRSCLKNYCEYEAKIMCSS-----ICKIACKNVEESPERKMLNSTPHY 205
DB 331 CNFTCEGFMLOG-----PAQVECTQGWTOQIIPCEAFQCTAL-SNFER-----CY 377
QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLRKNRQAFSCISWEVEATCACLLAQ-----253
DB 378 MNCLPSASGSFRYSGSCFSCGQFVLKGSKRLQCGPTGENDNEKPTCEAVRCDVHQPP 437
QY 254 -----GEEAEQHCPSLAPQMLIEFG-----RCLSQ 281
DB 438 KGLVRAHSPIGFTYKSKCAFSCEEG---FELYGSTQLECTSQ 478

RESULT 5

US-08-795-430-13

; Sequence 13, Application US/08795430

; Patent No. 6130071

; GENERAL INFORMATION:

; APPLICANT: Alitalo, Kari

;; APPLICANT: Joukov, Vladimir
;; TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)
;; NUMBER OF SEQUENCES: 57
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 6300 Sears Tower, 233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: United States of America
;; ZIP: 60606-6402
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/795,430
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/FI96/00427
;; FILING DATE: 01-AUG-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/671,573
;; FILING DATE: 28-JUN-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/601,132
;; FILING DATE: 14-FEB-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/585,895
;; FILING DATE: 12-JAN-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/510,133
;; FILING DATE: 01-AUG-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/340,011
;; FILING DATE: 14-NOV-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gass, David A.
;; REGISTRATION NUMBER: 38,153
;; REFERENCE/DOCKET NUMBER: 28967/33691
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312/474-6300
;; TELEFAX: 312/474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 13:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 418 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-795-430-13

Query Match 5.4%; Score 86; DB 4; Length 418;

Best Local Similarity 18.6%; Pred. No. 2;
Matches 48; Conservative 24; Mismatches 90; Indels 96; Gaps 10;

QY 14 CIDNCGARELKAHLHPQYDDQSFF-----POSELPKPMT-----48
DB 161 CGGCNCGSEGLQCMNISTNYISKTFLFTVPLSHGPKPVTVSFANHSTSCRCMSKLDYVRQV 220
QY 49 -TLVGRLLPVPAK-----LNLITQVDNGALPSAVNGAAPPSPALQ 88
DB 221 HSIIRSLPATQTCQHVANKTCPKNHVNWNNICRCLAQHDG-FSSHLGSDTSEGFHC 279
QY 89 GPPKITLSCYCDGCFSSGDF-----CN-----SCSC-----NLL-----116
DB 280 GPNKELDEETCCQCVCKGGVRRPISCGPHKELDRASCOCMKKLLPSSCGPNKEFDEKQ 339
QY 117 -----RHELEFRKAIKACLDNRNPEAFQPKMGKGRIGLAAKLRHSGKGNCKRSGLKN 167
DB 340 CVCKKTCPKHHPLNPAKCICECTESPNKCFL-----KG-----KRFHHQTCSCYRPPCTVR 390

QY 168 YCEYEAKIMCSSICKCI 185
Db 391 TKRCDAGFLIAEVCRCV 408

RESULT 6
US-09-355-700-13
; Sequence 13, Application US/09355700
; Patent No. 6361946
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research
; Helmsink University Licensing
; Alitalo, Kari (U.S. only)
; Joukov, Vladimir (U.S. only)
; TITLE OF INVENTION: Vascular Endothelial Growth Factor C (VEGF-C)
; Protein and Gene, Mutants Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/355,700
; FILING DATE: 05-NOV-1994
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/795,430
; FILING DATE: 05-FEB-1997
; APPLICATION NUMBER: PCT/F196/00427
; FILING DATE: 01-AUG-1996
; APPLICATION NUMBER: 08/671,573
; FILING DATE: 28-JUN-1996
; APPLICATION NUMBER: 08/601,132
; FILING DATE: 14-FEB-1996
; APPLICATION NUMBER: 08/585,895
; FILING DATE: 12-JAN-1996
; APPLICATION NUMBER: 08/510,133
; FILING DATE: 01-AUG-1995
; APPLICATION NUMBER: 08/340,011
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 28967/34140
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-355-700-13

Query Match 5.4%; Score 86; DB 4; Length 418;
Best Local Similarity 18.6%; Pred. No. 2;
Matches 48; Conservative 24; Mismatches 90; Indels 96; Gaps 10;

QY 14 CIDNCGARELKHLLPQYDQSF-----PQSELKPMTL-----VGR-LLPVA 59
Db 161 CGGCCNSEGLOCMNISTNYISKTILFTVPLSHGPKPVTVSFANHTSCRCMKLDVYRQV 220

QY 49 -TLVGRLLPVPK-----LNLITQVDNGALPSAVNGAAPSPGALQ 88
Db 221 HSIIRSLPATQTOCHVANKTCPKNHNWNNQICRLAQHDFG-FSSHLGSDISGEFHC 279

QY 89 GPKRITLSGYCDCFSSGDF-----CN-----SCSC-----NNL----- 116
Db 280 GPNKELDEETQCQVCKGKGVPRISGPHKELDRASCQCMCKNKLPLSSCGPNKFEDEKQ 339

QY 117 -----RHELEFRFAIKACLDNRNPEAFQPKMGKRLGAALKRHSKGCNKRSGCLN 167
Db 340 CVCKKTCPKHPLNPAKICICECTESPKNCFL-----KG-----KRFHQTCSCYRPPCTVR 390

QY 168 YCEYEAKIMCSSICKCI 185
Db 391 TKRCDAGFLIAEVCRCV 408

RESULT 7
5217870-2
; Patent No. 5217870
; APPLICANT: HESSON, CATHERINE A.; LOBB, ROY R.; GOELZ, SUSAN E.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST CDX
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/345,151
; FILING DATE: 28-APR-1989
; SEQ ID NO: 2;
; LENGTH: 610
5217870-2

Query Match 5.4%; Score 86; DB 6; Length 610;
Best Local Similarity 22.2%; Pred. No. 3.3;
Matches 64; Conservative 29; Mismatches 101; Indels 94; Gaps 17;

QY 15 IDNCGARE-----LKAHLHLLPQYDQSF-----SPQSELKPMTL-----VGR-LLPVA 59
Db 177 IVNCTALESPEHSGSLVCSHPLGNFYSNCSISCDRGYLPSSMETQMSSGWSADIPA 236

QY 60 ----KLNLIQVDNGALPSAVNGAAP-----SGPALQPPKAITLSGYCDCFSSG 105
Db 237 CNVVECDVATNPANGFVECFQNFQFMTCTFDCBEGFELMGAQSL-----OCTSSG 290

QY 106 DF-----CNSCSNNLRHELEFRFAIKACLDNRNPEAFQPKMGKRLG---AAKLHRSKG 156
Db 291 NWDNEKPTCKAVTCRAVR-----QPQNGSVRCSHSPAGEFTFKSS 330

QY 157 CN--CKRSGCLKNYCEYEAKIMCSS-----ICKCIACKNYEESPERKMLMSTPHY 205
Db 331 CNFTCEEGFMLOG-----PAQVECTTQGWTOQIPVCEAFQCTAL-SNPER-----GY 377

QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLKRNQAFSCISWEVVEATC 247
Db 378 MNCLPSASGSGFRYSGSCFESCEQGVFLKSKRLQCCGPTGEMDNEKPTC 425

RESULT 8
US-08-110-158-4
; Sequence 4, Application US/08110158
; Patent No. 5603821
; GENERAL INFORMATION:
; APPLICANT: McEver, Rodger P.
; APPLICANT: Pan, Junliang
; TITLE OF INVENTION: Expression Control Sequences of the
; TITLE OF INVENTION: P-Selectin Gene
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530


```

; REGISTRATION NUMBER: 31.815
; REFERENCE/DOCKET NUMBER: P31 8717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-728-215-31

Query Match 5.3%; Score 85.5; DB 2; Length 769;
Best Local Similarity 23.4%; Pred. No. 5;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps

QY 90 PPKITUSGYCDGFS-----SGD-----FCNSCSN-----NLRHELERFKAIKAC 129
Db 525 PGKLIYGYCEDTINCERYNGQVCGPGRGLFCGKCRCHPGFGSACQCR--TTEG 582
QY 130 LDNPENAFQPKWCKGRLGAALRHSGKCNCKRSGLKNKYCEYEAKIM-----CSSI 181
Db 583 L--NPRRVES-GRGR-----CRC-----NVCECHSGYQLPLCOECPCPCSP 621
QY 182 C----KCIACKNYEESP 194
Db 622 CGKYISCAECLNKEKGP 638

RESULT 13
US-08-938-085A-31
; Sequence 31, Application US/08938085A
; Patent No. 6339148
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,085A
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,215
; FILING DATE: 11-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
; REGISTRATION NUMBER: 42,058
; REFERENCE/DOCKET NUMBER: 023070-0802100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

```

```
; MOLECULE TYPE: protein
US-08-938-085A-31

Query Match          5.3%; Score 85.5; DB 4; Length 769;
Best Local Similarity 23.4%; Pred. No. 5;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps 10;

QY 90 PKKITLSGYCDGFS-----SGD-----FCNSCSN-----NLRHELEFRKAIRKAC 129
Db 525 PGKLIYQYCECDTINCERYNGVCGPGRLCGKCRCHPGFGSACQCR--TTEGC 582
QY 130 LDRNPEAFQPMGKGRGLGAALRHSGKNCGRSGCLKNVCEYEAKIM-----CSSI 181
Db 583 L--NPRRVECS-GRGR-----CRC-----NVCECHSGYQLPLCQCPGCPSP 621

QY 182 C-----KCIACKNYEESP 194
Db 622 CGKYISCAECLKFEKGP 638

RESULT 14
PCT-US95-04886-1
; Sequence 1, Application PC/TUS9504886
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..22
; OTHER INFORMATION: /label= signal
; FEATURE:
; NAME/KEY: Region
; LOCATION: 449..496
```

```
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 541..581
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 582..617
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 701..723
; OTHER INFORMATION: /label= trans
; OTHER INFORMATION: /note= "transmembrane domain"
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 724..769
; OTHER INFORMATION: /label= cyto
; OTHER INFORMATION: /note= "cytoplasmic domain"
; PUBLICATION INFORMATION:
; AUTHORS: Pigott,
; TITLE: LFA-1 Amino acid sequence (B2) (from human
; TITLE: tonsil)
; JOURNAL: The Adhesion Molecule Facts Book
; PAGES: 96-96
; DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
PCT-US95-04886-1

Query Match          5.3%; Score 85.5; DB 5; Length 769;
Best Local Similarity 23.4%; Pred. No. 5;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps 10;

QY 90 PKKITLSGYCDGFS-----SGD-----FCNSCSN-----NLRHELEFRKAIRKAC 129
Db 525 PGKLIYQYCECDTINCERYNGVCGPGRLCGKCRCHPGFGSACQCR--TTEGC 582
QY 130 LDRNPEAFQPMGKGRGLGAALRHSGKNCGRSGCLKNVCEYEAKIM-----CSSI 181
Db 583 L--NPRRVECS-GRGR-----CRC-----NVCECHSGYQLPLCQCPGCPSP 621

QY 182 C-----KCIACKNYEESP 194
Db 622 CGKYISCAECLKFEKGP 638

RESULT 15
PCT-US96-01314-45
; Sequence 45, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaout
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
```


Search completed: April 21, 2003, 11:51:09
Job time : 16.4091 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	396	24.8	438	10	US-09-220-091-13		Sequence 13, Appl
2	115	7.2	1192	9	US-10-189-971-18		Sequence 18, Appl
3	115	7.2	1207	9	US-10-189-971-20		Sequence 20, Appl
4	115	7.2	1477	9	US-10-189-971-8		Sequence 8, Appl
5	115	7.2	1535	9	US-10-189-971-14		Sequence 14, Appl
6	115	7.2	1593	9	US-10-189-971-4		Sequence 4, Appl
7	107.5	6.7	1057	9	US-10-189-971-6		Sequence 6, Appl
8	107.5	6.7	1251	9	US-10-189-971-16		Sequence 16, Appl
9	107.5	6.7	1342	9	US-10-189-971-24		Sequence 24, Appl
10	107.5	6.7	1512	9	US-10-189-971-10		Sequence 10, Appl
11	107.5	6.7	1570	9	US-10-189-971-12		Sequence 12, Appl
12	107.5	6.7	1628	9	US-10-189-971-2		Sequence 2, Appl
13	100	6.3	759	9	US-10-189-971-22		Sequence 22, Appl
14	97.5	6.1	652	10	US-09-789-919-96		Sequence 96, Appl
15	96	6.0	3635	9	US-10-037-182-4		Sequence 4, Appl
16	96	6.0	3635	10	US-09-845-583-2		Sequence 2, Appl
17	93	5.8	4123	9	US-10-213-509-5		Sequence 5, Appl
18	91.5	5.7	644	10	US-09-789-919-62		Sequence 62, Appl
19	91.5	5.7	644	10	US-09-789-919-73		Sequence 73, Appl

Db 367 EAATCMVAQAEEALNKEVQTEDEKLINMEKLVLEFRGRCLEQMI 412

RESULT 2

US-10-189-971-18
; Sequence 18, Application US/10189971
; Publication No. US20030028907A1

GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 18

; LENGTH: 1192

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-18

Query Match 7.2%; Score 115; DB 9; Length 1192;

Best Local Similarity 21.1%; Pred. No. 0.1;

Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGR 53

Db 352 VSEPKACAPALCPFPARGDCPCDCGCE-----YLGESYLSNQEFFPPREPCNLTCLGG 407

QY 54 LLPVPAKLNLTQVDNGALPSAVNGAAPPSPGALQGPPTITLSGYC-----DC-----F 102

Db 408 F-----VTCGRRPCEPPGCSHPLIP-----SGHCCTCGGCRYHGVTT 445

QY 103 SSGD-----FCNSCSCNNLRHLEFRFKAICACLDNPEAFQPKMKG-----GRL 146

Db 446 ASGETLPDPLDPTCSLCTCGREHQ-----DGEFEPGAGSCWCRCQAGQV 492

QY 147 GAAKLR-----HSKGCNCKR--SGCLK-----NYCECYE----- 173

Db 493 SCVRLQCPPLPCKLQVTERGSCCRGCLAHGEHEHPEGSRWVPPDSACSCVCHEGVTT 552

QY 174 -AKIMCSICK-----CIACKNYEESPERKMLMSTPHYMEPGD--FESSHYLSPAKF 222

Db 553 CARIOTISSCAQPROGPHDCPCDCGCE-----YEPGESFQPG--ADPCEV 601

QY 223 -----SGPPKLRNQAFCISWEVVEATCACLLAQGEAEQEHCSPSLAQEM 270

Db 602 CICEPQEPGPPSLRCHRR--QCPS--LVGCPPSLLPPG-----PQHCCPTCAEAL 648

RESULT 3

US-10-189-971-20

; Sequence 20, Application US/10189971

; Publication No. US20030028907A1

GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 20

; LENGTH: 1192

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-18

; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20

; LENGTH: 1207

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-20

Query Match 7.2%; Score 115; DB 9; Length 1207;

Best Local Similarity 21.1%; Pred. No. 0.1;

Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGR 53

Db 367 VSEPKACAPALCPFPARGDCPCDCGCE-----YLGESYLSNQEFFPPREPCNLTCLGG 422

QY 54 LLPVPAKLNLTQVDNGALPSAVNGAAPPSPGALQGPPTITLSGYC-----DC-----F 102

Db 423 F-----VTCGRRPCEPPGCSHPLIP-----SGHCCTCGGCRYHGVTT 460

QY 103 SSGD-----FCNSCSCNNLRHLEFRFKAICACLDNPEAFQPKMKG-----GRL 146

Db 461 ASGETLPDPLDPTCSLCTCGREHQ-----DGEFEPGAGSCWCRCQAGQV 507

QY 147 GAAKLR-----HSKGCNCKR--SGCLK-----NYCECYE----- 173

Db 508 SCVRLQCPPLPCKLQVTERGSCCRGCLAHGEHEHPEGSRWVPPDSACSCVCHEGVTT 567

QY 174 -AKIMCSICK-----CIACKNYEESPERKMLMSTPHYMEPGD--FESSHYLSPAKF 222

Db 568 CARIOTISSCAQPROGPHDCPCDCGCE-----YEPGESFQPG--ADPCEV 616

QY 223 -----SGPPKLRNQAFCISWEVVEATCACLLAQGEAEQEHCSPSLAQEM 270

Db 617 CICEPQEPGPPSLRCHRR--QCPS--LVGCPPSLLPPG-----PQHCCPTCAEAL 663

RESULT 4

US-10-189-971-8

; Sequence 8, Application US/10189971

; Publication No. US20030028907A1

GENERAL INFORMATION:

; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 1477

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-8

Query Match 7.2%; Score 115; DB 9; Length 1477;

Best Local Similarity 21.1%; Pred. No. 0.13;

Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

QY 2 VICOLKGAQML-----CIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGR 53

Db 637 VSEPKACAPALCPFPARGDCPCDCGCE-----YLGESYLSNQEFFPPREPCNLTCLGG 692

QY 54 LLPVPAKLNLTQVDNGALPSAVNGAAPPSPGALQGPPTITLSGYC-----DC-----F 102

Db 693 F-----VTCGRRRPCEPPGCSHPLIP-----SGHCCPTCQGCYHGVTT 730
Qy 103 SSGD-----FCNSCSCNNLRHELFRFAKACLDNPEAFQPKMGK-----GRL 146
Db 731 ASGETLPDPLDPTCSLCTCQGREHQ-----DGEFEGPAGSCWCRCQAGQV 777
Qy 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCECYE----- 173
Db 778 SCVRLQCPPLPKLQVTERGSCPCRCGLAHGEEHPEGSRWVPDSDACSSVCHEGVVT 837
Qy 174 -AKTMCSSICK-----CIACKNYEESPERKMLMSTPHYMEPGD-FESSHYLSPAKF 222
Db 838 CARIQCISSCAQPRGPHDCCPQCSDCEHEGRK-----YEPGESFQPG--ADPCEV 886
Qy 223 -----SGPPKLRNKAQAFSCISWEVVEATCACLAAQGEAEQEHCSPSLAEOM 270
Db 887 CICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----POHCCPTCAEAL 933

RESULT 5
US-10-189-971-14
; Sequence 14, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1535
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-14

Query Match 7.28; Score 115; DB 9; Length 1535;
Best Local Similarity 21.18; Pred. No. 0.14;
Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

Qy 2 VICOLKGAQML-----CIDMCGARELKALHLLPOYDQSSFPQSELKPKMTTLVGR 53
Db 695 VSCEPKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPCNLTCLGG 750
Qy 54 LLPVPAKLNLTQVNDGALPSAVNGAAPSGPALQGPKITLSGYC-----DC-----F 102
Db 751 F-----VTCGRRRPCEPPGCSHPLIP-----SGHCCPTCQGCYHGVTT 788
Qy 103 SSGD-----FCNSCSCNNLRHELFRFAKACLDNPEAFQPKMGK-----GRL 146
Db 789 ASGETLPDPLDPTCSLCTCQGREHQ-----DGEFEGPAGSCWCRCQAGQV 835
Qy 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCECYE----- 173
Db 836 SCVRLQCPPLPKLQVTERGSCPCRCGLAHGEEHPEGSRWVPDSDACSSVCHEGVVT 895
Qy 174 -AKTMCSSICK-----CIACKNYEESPERKMLMSTPHYMEPGD-FESSHYLSPAKF 222
Db 896 CARIQCISSCAQPRGPHDCCPQCSDCEHEGRK-----YEPGESFQPG--ADPCEV 944
Qy 223 -----SGPPKLRNKAQAFSCISWEVVEATCACLAAQGEAEQEHCSPSLAEOM 270
Db 945 CICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----POHCCPTCAEAL 991

RESULT 6
US-10-189-971-4
; Sequence 4, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-4

Query Match 7.28; Score 115; DB 9; Length 1593;
Best Local Similarity 21.18; Pred. No. 0.14;
Matches 75; Conservative 34; Mismatches 102; Indels 144; Gaps 21;

Qy 2 VICOLKGAQML-----CIDMCGARELKALHLLPOYDQSSFPQSELKPKMTTLVGR 53
Db 753 VSCEPKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPCNLTCLGG 808
Qy 54 LLPVPAKLNLTQVNDGALPSAVNGAAPSGPALQGPKITLSGYC-----DC-----F 102
Db 809 F-----VTCGRRRPCEPPGCSHPLIP-----SGHCCPTCQGCYHGVTT 846
Qy 103 SSGD-----FCNSCSCNNLRHELFRFAKACLDNPEAFQPKMGK-----GRL 146
Db 847 ASGETLPDPLDPTCSLCTCQGREHQ-----DGEFEGPAGSCWCRCQAGQV 893
Qy 147 GAAKLR-----HSKGCNKR-SGCLK-----NYCECYE----- 173
Db 894 SCVRLQCPPLPKLQVTERGSCPCRCGLAHGEEHPEGSRWVPDSDACSSVCHEGVVT 953
Qy 174 -AKTMCSSICK-----CIACKNYEESPERKMLMSTPHYMEPGD-FESSHYLSPAKF 222
Db 954 CARIQCISSCAQPRGPHDCCPQCSDCEHEGRK-----YEPGESFQPG--ADPCEV 1002
Qy 223 -----SGPPKLRNKAQAFSCISWEVVEATCACLAAQGEAEQEHCSPSLAEOM 270
Db 1003 CICEPQEGPPSLRCHRR--QCPS--LVGCPPSOLLPPG-----POHCCPTCAEAL 1049

RESULT 7
US-10-189-971-6
; Sequence 6, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 4.0

```

; SEQ ID NO 6
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-6

Query Match
Best Local Similarity 6.7%; Score 107.5; DB 9; Length 1057;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICQLKGGQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMTTLVGR 53
Db 182 VSCPRKACAPALCPFPARGDCCPDGCE-----YLGESYLSNOEFFDPPREPCNLCTCLGG 237
QY 54 LLP-----VPA-----KLNLTQVDNGALPSAVNGAAF----- 81
Db 238 FVTGRRPCEPPGCSHPLPSGHCCPTCGCRYHGVTTASGETLPDLPDPTCSLCTCQEG 297
QY 82 -----PSGPALQGPCKITLSGYCDCFSSGDFCNSCSCNNLRHELEKFAIKACLD RNP 134
Db 298 SMRCQKKPCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 337
QY 135 EAFQPKMGK-----GRGAAKLR-----HSKGCNCKR-SGCLK----- 166
Db 338 EEFEGPAGSCWCRCQAGVSCVRLQCPPLCKLQVTERGSCCPCRCGCLAHGEEHPEGS 397
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
Db 398 RWVPPDSACSSCVCHGVVTCARIQISSCAQPROGPHDCCPQCSDCHEGRK----- 450
QY 204 HYMEPGD-FESSHYLSPAF-----SGPPKLRKQAFSCISWEVVEATCACLIAQGE 255
Db 451 --YEPGESFQPG--ADPCEVCICEPOEPGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 501
QY 256 EAEQEHCSPLABQM 270
Db 502 ---PQHCCPTCAEAL 513

RESULT 8
US-10-189-971-16
; Sequence 16, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleoti
; FILE REFERENCE: Same
; CURRENT APPLICATION NUMBER: US/10/189,971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-16

Query Match
Best Local Similarity 6.7%; Score 107.5; DB 9; Length 1251;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICQLKGGQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMTTLVGR 53
Db 376 VSCPRKACAPALCPFPARGDCCPDGCE-----YLGESYLSNOEFFDPPREPCNLCTCLGG 431
QY 54 LLP-----VPA-----KLNLTQVDNGALPSAVNGAAF----- 81
```

```

Db 432 FVTGRRPCEPPGCSHPLPSGHCCPTCGCRYHGVTTASGETLPDLPDPTCSLCTCQEG 491
QY 82 -----PSGPALQGPCKITLSGYCDCFSSGDFCNSCSCNNLRHELEKFAIKACLD RNP 134
Db 492 SMRCQKKPCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 531
QY 135 EAFQPKMGK-----GRGAAKLR-----HSKGCNCKR-SGCLK----- 166
Db 532 EEFEGPAGSCWCRCQAGVSCVRLQCPPLCKLQVTERGSCCPCRCGCLAHGEEHPEGS 591
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
Db 592 RWVPPDSACSSCVCHGVVTCARIQISSCAQPROGPHDCCPQCSDCHEGRK----- 644
QY 204 HYMEPGD-FESSHYLSPAF-----SGPPKLRKQAFSCISWEVVEATCACLIAQGE 255
Db 645 --YEPGESFQPG--ADPCEVCICEPOEPGPPSLRCHRR--QCPS--LVGCPPSOLLPPG- 695
QY 256 EAEQEHCSPLABQM 270
Db 696 ---PQHCCPTCAEAL 707

RESULT 9
US-10-189-971-24
; Sequence 24, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 1342
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-24

Query Match
Best Local Similarity 6.7%; Score 107.5; DB 9; Length 1342;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

QY 2 VICQLKGGQML-----CIDNCGARELKALHLLPOYDDQSSFPQSELPKPMTTLVGR 53
Db 467 VSCPRKACAPALCPFPARGDCCPDGCE-----YLGESYLSNOEFFDPPREPCNLCTCLGG 522
QY 54 LLP-----VPA-----KLNLTQVDNGALPSAVNGAAF----- 81
Db 523 FVTGRRPCEPPGCSHPLPSGHCCPTCGCRYHGVTTASGETLPDLPDPTCSLCTCQEG 582
QY 82 -----PSGPALQGPCKITLSGYCDCFSSGDFCNSCSCNNLRHELEKFAIKACLD RNP 134
Db 583 SMRCQKKPCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 622
QY 135 EAFQPKMGK-----GRGAAKLR-----HSKGCNCKR-SGCLK----- 166
Db 623 EEFEGPAGSCWCRCQAGVSCVRLQCPPLCKLQVTERGSCCPCRCGCLAHGEEHPEGS 682
QY 167 -----NYCEYE-----AKIMCSSICK-----CIACKNYEESPERKMLMSTP 203
Db 683 RWVPPDSACSSCVCHGVVTCARIQISSCAQPROGPHDCCPQCSDCHEGRK----- 735
```



```
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-2

Query Match
Best Local Similarity 6.7%; Score 107.5; DB 9; Length 1628;
Matches 75; Conservative 37; Mismatches 114; Indels 149; Gaps 21;

Qy 2 VICOLKGAOML-----CIDNCGARELKAHLPOYDDQSSGFPQSELKPKMTTLVGR 53
Db 753 VSCEPKACAPALCPFPARGDCPCDGC-----YLGESYLSNQEFPPDPREPCNLTCLGG 808
Qy 54 LLP-----VPA-----KLNLIQVDNGALPSAVNGAAF-----81
Db 809 FVTCGRRRCEPPGSHPLIPSGHCHCPTCQCGRYHGVTTASGETLPDPLDPTCSLCTCQEG 868
Qy 82 -----PSGALOGPPKITLSGYCDGFCSSGDFCNSCNLRHLEFRFAKACLDNRN 134
Db 869 SMRCQKRCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DG 908
Qy 135 EAFQPKMGK-----GRLGAALKR-----HSKGCNCKR-SGCLK-----166
Db 909 BEFEGPAGCWCRCQAGQVSCVRLQCPKLVQTERGSCCPCRCGLAHGEHPGGS 968
Qy 167 -----NYCECYE-----AKIMCSICK-----CIACKNYEESPERKMLMSTP 203
Db 969 RWPVPSDSCSCVCHGVVTCARTICISSCAQPRGPHDCCPQSCDCEHGRK-----1021
Qy 204 HMEPGD-FESHYLSPAKF-----SGPKLKRNRQAFSCISWEVVEATCACLLAQGE 255
Db 1022 --YEPGESFQPG--ADPCEVCICEPQEPGPPSLRCHRR--QCPS--LVGCPSPQLPPG- 1072
Qy 256 EAEQEHCSPSLAEQM 270
Db 1073 ---POHCPTCAEAL 1084

RESULT 13
US-10-189-971-22
; Sequence 22, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kielin-like Proteins and Polynucleotides
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; PRIOR FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 759
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-22

Query Match
Best Local Similarity 6.3%; Score 100; DB 9; Length 759;
Matches 56; Conservative 24; Mismatches 70; Indels 98; Gaps 16;

Qy 82 PSGALOGPPKITLSGYCDGFCSSGDFCNSCNLRHLEFRFAKACLDNRNPEAFQPKM 141
Db 7 PCAPALCPHPS---PGPCFC---PVCHSCLSQGREHQ-----DGEFEFGPA 46
Qy 142 GK-----GRLGAALKR-----HSKGCNCKR-SGCLK-----166
Db 47 GSCWCRCQAGQVSCVRLQCPKLVQTERGSCCPCRCGLAHGEHPGSRWVPPDS 106
```

```
Qy 167 --NYCECYE-----AKIMCSICK-----CIACKNYEESPERKMLMSTPHYMPGD 210
Db 107 ACSSCVCHGVVTCARTICISSCAQPRGPHDCCPQSCDCEHGRK-----YEPGE 157
Qy 211 -FESHYLSPAKF-----SGPKLKRNRQAFSCISWEVVEATCACLLAQGEAEQEH 262
Db 158 SFQPG--ADPCEVCICEPQEPGPPSLRCHRR--QCPS--LVGCPSPQLPPG-----PQHC 207
Qy 263 SPSLAEQM 270
Db 208 CPTCAEAL 215
```

```
RESULT 14
US-09-789-919-96
; Sequence 96, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor
; APPLICANT: Moore, Kateri
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM
; TITLE OF INVENTION: CELLS AND USES THEREOF
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 96
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-96
```

```
Query Match
Best Local Similarity 6.1%; Score 97.5; DB 10; Length 652;
Matches 31; Conservative 12; Mismatches 49; Indels 19; Gaps 4;

Qy 96 SGYC--DCFSSGDFCNSCNSNLRHLEFRFAKACLDNRNPEAFQPKMGRLGAALKRH 153
Db 268 NGGCHQDCFEFGDGSFLCGCRPRRLD---DLVTCASRNPCSSSPCRG-----313
Qy 154 SKGCNCKRSGCLKNY-CEYEAKIMCSICKYACIKACNYEESPERKMLMSTP 203
Db 314 --GATCVLGHGKNYKTCRCPOGYQLDSSQLDCVDVDECQDSPCAQECVNTP 362
```

```
RESULT 15
US-10-037-182-4
; Sequence 4, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thybrell, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3635
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-182-4
```

```
Query Match
Best Local Similarity 6.0%; Score 96; DB 9; Length 3635;
Matches 18.0%; Pred. No. 18;
```



```
Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;
QY 14 CIDNCGARELKALHLLPOYDDOSSPQSELPKPMTLYGRLLLPVPAKLNLIITQVDNGALP 73
Db 117 CLERFGPRTLER---ITODDDVI-----CTEYSRIVPL-ENGEIVVSLVNGR-P 161
QY 74 SAVNGAAFP-----SGPALQGP-----PKITLSGY 98
Db 162 GAINFSYSPLLRDFTKATNIRLFLRTNTLLGLHMGKALRDPVTYRRYYYSIKDISIGR 221
QY 99 CDCFSSGDFC-----NSC-----SC 113
Db 222 CYCHGHADYCDAKDPLDPFLQACQHNCTCGGSCDRCCPGFNQOPWKPATTTDSANECQSC 281
QY 114 NNLRH-----ELERFKAKACLDRNPEAFQPKMGKRLGAAL---RHSKGCNCKRS 162
Db 282 NCHGHAYDCYDPEVDR-----RNASQNDNVYOG--GGVCLDCQHHHTGINCER- 329
QY 163 GCLKNY-----CECYEAKI---MCSSI-----CK-----CIAC-KNY 190
Db 330 -CLPGFFRAPDQPLDSPHVCRPCDC-ESDFTDGTCTEDLTGRCYCRPNFTGELCAACABGY 387
QY 191 EESPERKMLMSTPHMEPGDPESHYLSPAKESGPPKLRKNRQAFSCISWEVVEATCACL 250
Db 388 TDFPHCYPLPSFPHN-----DTREQVLPA-----GQIVNCDCNAA 422
QY 251 LAQGEAEQE-----HC-----SPSLAEQMILEEFG 277
Db 423 GTQGNACRKDPRLGRVCVKPNFRGAHCELCAFGFHGSPCHPCQCSSPGVANSLCDPESGQ 482
QY 278 CL 279
Db 483 CM 484
```

Search completed: April 21, 2003, 11:51:47
Job time : 19.4024 secs

1	1599	100.0	295	21	US-09-743-237-4	Sequence 4, Appl
2	1599	100.0	295	21	US-09-743-237-24	Sequence 24, Appl
3	1251	78.2	299	21	US-09-743-237-5	Sequence 5, Appl
4	505.5	31.6	223	1	PCT-US01-08656-10693	Sequence 10693, A
5	435	27.2	403	26	US-60-161-932-1700	Sequence 1700, Ap
6	435	27.2	890	26	US-60-167-217-12948	Sequence 12948, A
7	435	27.2	950	20	US-09-614-150-12897	Sequence 12897, A

```

Query Match      100.0%; Score 1599; DB 21; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.3e-143;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MYICOLKGAOMLCIDNCARELKALHLLPQYDDQSSFQSELPKPMTTVLGRLLPYPAK 60
db
1 MYICOLKGAOMLCIDNCARELKALHLLPQYDDQSSFQSELPKPMTTVLGRLLPYPAK 60

```



```

Db      58 ASRPKPCNCTKSCLCLKLYCDCFANGEFONCNCNTCYNNLEHENEQRKAICACLDNRPE 117
Qy      136 AFOPKMKGRLGAALRHRSKGCNCRSGCLKNYCEYEAKIMCSSICKCIACKNYEESPE 195
       ||::|||:: | : ||||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      118 AFRPKIGKGESEDRHSKGCNCRSGCLKNYCEYEAKIMCSSICKCIACKNYEESPE 177
Qy      196 RKMLM 200
       || ||
Db      178 RKTLM 182

RESULT      5
US-60-161-932-1700
; Sequence 1700, Application US/60161932
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic
; TITLE OF INVENTION: Acid Sequences, Systems Containin the Nucleic Acid Sequences
; FILE REFERENCE: CL000122
; CURRENT FILING DATE: 1999-11-28
; NUMBER OF SEQ ID NOS: 2626
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 1700
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Drosophila
US-60-161-932-1700

Query Match          27.2%; Score 435; DB 26; Length 403;
Best Local Similarity 38.4%; Pred No. 2.6e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

Qy      31 QYDDQSFFPOSELPKMTLLVGRLLPVPAPKLNLITQVDNGALPSAVNGAAFPSPALQGP 90
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      134 QLPTQEQTPIKVEPKLPTLPVGKANVPAPK--PLFEVLKPPTATAAAGAVDPLGGMTSRR 191
Qy      91 PKITLSG-----YDCFFSGDFCNCSGC-----NNLRHELERFAIKACLDNRPEAFQPK 140
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      192 KHCNCKSQCLKLYCCFCANGEEQCQDCTCKCFNNLDYEVERERAIRSCLDNRNPFAFKPK 251
Qy      141 MGKGRGLAAALRHRSKGCNCRSGCLKNYCEYEAKIMCSSICKIACKNYESPERRKMLM 200
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      252 ITAPNSGDML--HNKCGNCRSGCLKNYCEYEAKIPCSICKVCGRNMEDRPD----- 305
Qy      201 STHYMPEGDFESSHYLSPAKFSGPPKLRNKRO-----AFSCISWEWBEATCACILLAQGE 255
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      306 -----VDMDSLGLMGVEGQKKDK-AKNKQLNENRANIVFTDDVIEATIMCMISRIV 356
       | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      256 EAQEHCSPSLAEQMLEEPGRCLSQL 283
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      357 MHEKQNVAVEDMEREVMEENGESLTQII 384

RESULT      6
US-60-167-217-12948
; Sequence 12948, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 12948
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Drosophila

```



```

PCT-US01-086311-48317
; Sequence 48317, Application PC/TUSO108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/USO1/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 48317
; LENGTH: 280
; TYPE: PRF
; ORGANISM: Homo sapiens
PCT-US01-086311-48317

Query Match          22.6%; Score 362; DB 1; Length 280;
Best Local Similarity 57.9%; Pred. No. 1.5e-25;
Matches 66; Conservative 15; Mismatches 23; Indels 10; Gaps 2;

QY   98 YCDDCFSGDFCNCSGC-----NNLRHRLERKAIACLDNRNEAFQPKMGKGRLGAAKLURH 153
      |||||:::||||::||    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db   114 YCDGFANGFCNCCNCNTCYNNLNHEHERQAIAKLDRNEAPKPICKGKEGESDRRH 173
            :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY   154 SKGNCKRSGCLKNYCCEYAEIKMSSICKCIACKNYEESPERKMLMSTPHYME 207
      |||||||::::::::::||  ::  :  :  :  :  :  :  :  :  :  :  :  :  :
Db   174 SKGNCKRSGCLKNYCCEYENIIFWCCR-----YRQLKAQLLSGAFLGLE 221


RESULT 15
US-09-708-427-33367
; Sequence 33367, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLY
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33367
; LENGTH: 524
; TYPE: PRF
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..524
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..524
; OTHER INFORMATION: Ceres Seq. ID 1836306
US-09-708-427-33367

Query Match          22.3%; Score 357; DB 21; Length 524;
Best Local Similarity 32.2%; Pred. No. 1e-24;
Matches 79; Conservative 39; Mismatches 73; Indels 54; Gaps 8;

QY   41 SELPKPMWTTLVGRL--PVPAAKINLLITQVD-NGALPSAVNGAAPPSPGPA 86
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db   4 ASIPSPIVTTRPIITSQAAPTATPPPPPQSOGIIHLHPIRPRPEPNMPRPAGET 63

QY   87 LOGPPK-----ITLSGYCDCFFSGDCFNCSGC-----NLRLHELRFKATKACLDR 132
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db   64 RDGPQPQKKQCOKNHKSRLCLLYECFAFGSYTCDCGCCNVCMNFENNBPBPARQVAESTLER 123

QY   133 NPFAFPQPMKGKRGLCAAK-----LRHSKGCNCRSGLCKLKNYCECYEAkimcSSI 181
```


Query Match 12.8%; Score 205; DB 6; Length 359;
Best Local Similarity 24.5%; Pred. NO. 1.7e-10;
Matches 64; Conservative 31; Mismatches 90; Indels 76; Gaps 10;

```

; NUMBER OF SEQ ID NOS: 1652
;
; SOFTWARE: FastSEQ for Windows version 4.0

```

SEQ ID NO 1127
LENGTH: 120

TYPE: PRT
ORGANISM: Lolium perenne
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(120)
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-138-145-1127

Query Match 9.9%; Score 159; DB 6; Length 120;

Best Local Similarity 38.7%; Pred. No. 5.9e-07;
Matches 41; Conservative 13; Mismatches 34; Indels 18; Gaps 7;

QY 98 YDCFFSGDFCNS-CSCN---NLRHELERFKAIAKACLDNRPEAFQPMKGRGLGA-----148

Db 16 YCEC-AAGVYCEPCSCGCLNKPHEIIVLSTRKQIERNPLA-APKVRILSDAAQETQE 73

QY 149 -----AKLRHSGKNCNRSGCLNKCCEYAKIMCISICKIACKN 189

Db 74 DPNTPASARHKRCNCKKSLK-YCECYQGGVGLTNCR-ECNK 117

RESULT 10

US-10-155-881-9182

Sequence 9182, Application US/10155881

GENERAL INFORMATION:

APPLICANT: Dotson, Stanton B.

APPLICANT: Kovalic, David K.

APPLICANT: Liu, Jingdong

APPLICANT: Lutfiyya, Linda L.

APPLICANT: McIninch, James

TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

TITLE OF INVENTION: TRANSCRIPTION IN PLANTS

FILE REFERENCE: 38-21(15300)J

CURRENT APPLICATION NUMBER: US/10/155,881

CURRENT FILING DATE: 2002-05-22

NUMBER OF SEQ ID NOS: 37595

SEQ ID NO 9182

LENGTH: 200

TYPE: PRT

ORGANISM: Zea mays

US-10-155-881-9182

Query Match

Best Local Similarity 24.7%; Pred. No. 0.0056;

Matches 39; Conservative 15; Mismatches 57; Indels 47; Gaps 4;

QY 39 PQSELKPMTLVG-----RLPVP--AKNLITQVDNGALPSAVNGAAPPSPGALQGGP 91

Db 26 PQAAVPMDADYGAGATNPWPEPDHLSPPPPPPASAPTPAAEAAPTASAAVSPP 85

QY 92 KITLGVDCDFSSGDFCNSCSNNLRHELERFKAIAKACLDNRPEAFQPMKGRGLGA 151

Db 86 -----APRANGNSDKRRRRADDDG---107

QY 152 RHBKGNCRSGCLNKCCEYAKIMCISICKIACKN 189

Db 108 --CKACSKRCKLKLYCYVYASGSHCTELCGCEPCHN 143

RESULT 11

PCT-US02-01339-6

Sequence 6, Application PC/TUS0201339

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.

APPLICANT: LEE, Ernestine A.

APPLICANT: WALIA, Narinder K.

APPLICANT: BAUGHN, Mariah R.

APPLICANT: AZIMZAI, Yalda

APPLICANT: TANG, Y. Tom

APPLICANT: YUE, Henry
APPLICANT: THANGAVELU, Kavitha
APPLICANT: XU, Yuming
APPLICANT: ARVIZU, Chandra
APPLICANT: WARREN, Bridget A.
APPLICANT: YAO, Monique G.
APPLICANT: AU-YOUNG, Janice
APPLICANT: HAFALIA, April J.A.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: KALLICK, Deborah A.
APPLICANT: GANDHI, Ameena R.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: KHAN, Farrah A.
APPLICANT: LU, Yan
APPLICANT: SWARNAKAR, Anita
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: NGUYEN, Dannie B.
APPLICANT: GRAUL, Richard
APPLICANT: LU, Dyung Alina M.
TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
FILE REFERENCE: PI-0346 PCT
CURRENT APPLICATION NUMBER: PCT/US02/01339
CURRENT FILING DATE: 2002-01-16
PRIOR APPLICATION NUMBER: 60/262,838; 60/265,927; 60/271,196; 60/274,549; 60/334,1
PRIOR FILING DATE: 2001-01-19; 2001-02-02; 2001-02-23; 2001-03-09; 2001-11-28
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PERL Program
SEQ ID NO 6
LENGTH: 1774
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7722591CDI
PCT-US02-01339-6

Query Match 7.2%; Score 115; DB 1; Length 1774;

Best Local Similarity 23.1%; Pred. No. 0.13;

Matches 73; Conservative 27; Mismatches 124; Indels 92; Gaps 18;

QY 17 NCGARELKALHLLPQYDDQSSPQSELKPMTLVGRLLPVPKAKNLITQVDNGALPSAV 76

Db 748 NCNA-----WLLPDLTRSAS-----VGPMEESVAHAVAAGSRLYISGGFGVALGR 796

QY 77 NGAAPFSGPA--LQGPKITLSGYCD-CFS---SGD-----FCNSCSNNL---RHELER 122

Db 797 LALTPPDCRLLSPEACNQSGACTWCHGACLSGDAHRLCGGSGPCSPMPSPECRR 856

QY 123 FRAIKACLDNRPEAFQPMKGRGLGAALKRHSKGC-NCKRSGL-----KNYC-----169

Db 857 LRTCSCLARHPRTLQPGDGE-----ASTPRCKWCTNCPEGACIGRNGSCTSENDCRINQ 911

QY 170 -----ECYEAKI-----MCSSICKIACKNYEESPE-RKMLMSTPHY-----205

Db 912 REVFWAGNCSAAGCAACDCEQCTRECKCMWTRQFKRTGETRRLSVQPTDTCFSSHLL 971

QY 206 -MEPGDFESSHYLSPAKFSGPPKLRKNROAFSCISWEVVEATCACLLAOGAEAEQHCSP 264

Db 972 NYSPMPVSS---PPLPCTPTCHLLPN-----CTS-----CLDSKAGGGHQHCVM 1014

QY 265 SLAEQMILEEFGRCLS 280

Db 1015 SSSLQ-----CQLS 1023

RESULT 12

PCT-US02-10812-17

Sequence 17, Application PC/TUS0210812

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.

APPLICANT: GANDHI, Ameena R.

APPLICANT: DELEGEANE, Angelo M.

```

Query Match      5.9%; Score 94.5; DB 1; Length 538;
Best Local Similarity 20.5%; Pred. No. 2.1;
Matches 62; Conservative 37; Mismatches 110; Indels 93; Gaps 17;

QY 14 CIDNGARELKALHLLPOYDDQSSFOSELPKPMTTLVGRLLPVPAPAKLNLITQVNDGALP 73
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 76 CYSNCSPPKFQILML-----FPPNLYKPEIT-----LEAFAVITQM----- 112
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 74 SAVNGAAPSPGALQPPKITLSGYCOCFFSSGFCN-----SCSCNNLRHEL 120
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 113 -----LALSIGISYDDPRK-----CQSESTCINPEVQSGVKFTSSCSLSRFQFI 161
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 121 ERFKAIRACLDRNPEAFQPKMGKGLGAALKRHSGKNCK-----RSGCLKN 167
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 162 SNV-GVR-CLONKPO-HQKSKPKVCGNGRLEGNICDCGTEAOGGPASCCDFTVCULKD 218
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 168 YCECYEAKIMCSSTCK-----IACKNYEESPERKMLMSTPHMEPGDFESSHLYSPAKFS 223
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 219 GAKCY--KGLCKKDCQILQSGVECRP-KAHPECDIA-----ENCGSSPE--C 261
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
QY 224 GPP-----KLKNRQAFCSTISWEV--VEATCACLLAGEEAEQEHCSPLAEQMILEBFG 276
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :
Db 262 GPDTTLINGLSCKNNKFCYDGDCHDLDAECSEVFGSRNAPACPYEEIOSQS--DRFG 319
      | | | : | | | : | | | : | | | : | | | : | | | : | | | :

```

[illegible]

Db 113 -----LALSIGISYDDPK-----CQSESTCIMPEVVOVSGVKTFFSSCLSRSTQNF 161

QY 121 ERFAKAIACLDNRPEAFQPMGKGRGLGAALKRHSKCCKCK-----RSGCLKN 167

Db 162 SNV-GVK-CLQNKPO-MQKSPKPCVCGNGRLEGNEICDCGTEAOCGPASCDFRTCVLKD 218

QY 168 YCEYEAKIMCSICKC-----TACKNYESPSPKMLMSTPHMYEPDGFESSHYLSPAKFS 223

Db 219 GAKCY--KGLCKDCQILQSGVECRP-KAHPECIDIA-----ENCGSSPE--C 261

QY 224 GPP-----KLKRNQAFSGISMEV--VEATCACLLAQBEAEQEHGSPSLAEQMILEEFG 276

Db 262 GPDITILGLSCKNNKFCYDGDCHDLARCESVFGKGRNAPFACYEIIQSOS--DREG 319

QY 277 RC 278

Db 320 NC 321

RESULT 14

US-10-155-881-33654

; Sequence 33654, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH TRANSCRIPTION IN PLANTS

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 33654

; LENGTH: 298

; TYPE: PR1

; ORGANISM: Zea mays

US-10-155-881-33654

Query Match 5.8%; Score 92.5; DB 6; Length 298;

Best Local Similarity 23.7%; Pred. No. 1.6;

Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;

QY 94 TLSGYDCDFSSGDFCNS-CSCNNLRHELRFKAIKACLDNRPEAFQPMGKGRGLGAALKR 152

Db 53 TCGKQCPCLSNCTCCEKYGC-----PKICKNRF----- 81

QY 153 HSKGCKRSGCLKNYCEYEAKIMCS-SICK-C-IAC-----KNYE 191

Db 82 --RGCHCAKSQCRSQCPQCFADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 133.

RESULT 15

US-10-155-881-11197

; Sequence 11197, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH TRANSCRIPTION IN PLANTS

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 11197

; LENGTH: 580

; TYPE: PR1

; ORGANISM: Zea mays

US-10-155-881-11197

Query Match 5.8%; Score 92.5; DB 6; Length 580;

Best Local Similarity 23.7%; Pred. No. 3.5;

Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;

QY 94 TLSGYDCDFSSGDFCNS-CSCNNLRHELRFKAIKACLDNRPEAFQPMGKGRGLGAALKR 152

Db 335 TCGKQCPCLSNCTCCEKYGC-----PKICKNRF----- 363

QY 153 HSKGCKRSGCLKNYCEYEAKIMCS-SICK-C-IAC-----KNYE 191

Db 364 --RGCHCAKSQCRSQCPQCFADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 415

Search completed: July 11, 2002, 08:14:15

Job time: 39 sec

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:45:13 ; Search time 15.3956 Seconds
(without alignments)
1842.060 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLIHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	396	24.8	429	2 T23152	hypothetical prote
2	363	22.7	571	2 B84585	hypothetical prote
3	357	22.3	603	2 T08955	hypothetical prote
4	232	14.5	658	2 F71410	hypothetical prote
5	115	7.2	1737	2 T00209	MEGF8 protein - hu
6	100	6.3	775	2 T21436	hypothetical prote
7	97	6.1	497	2 T27827	hypothetical prote
8	96.5	6.0	1790	1 MMFR81	laminin beta-1 cha
9	96	6.0	769	2 S55524	male-specific leth
10	96	6.0	3635	2 T10053	laminin alpha 5 ch
11	95	5.9	1096	1 S61917	protein kinase C (
12	95	5.9	1428	2 T08852	lustrin A - Califo
13	94	5.9	716	2 T26998	hypothetical prote
14	94	5.9	1819	2 A71928	cag island protein
15	93	5.8	1816	1 S68960	laminin alpha-4 ch
16	92	5.8	798	2 A28193	integrin beta-1 ch
17	91.5	5.7	383	2 D88633	protein F56B3.2 [i
18	91	5.7	1927	2 G54585	cag pathogenicity
19	91	5.7	2907	2 A57278	fibrillin-2 precur
20	90.5	5.7	735	2 G02937	fertilin beta - cr
21	89	5.6	574	2 B88465	protein B0244.8 [i
22	89	5.6	1786	1 MMWSB1	laminin beta-1 cha
23	88	5.5	625	2 F4706	probable vacuolar
24	88	5.5	798	2 B28193	integrin beta-1* c
25	88	5.5	1786	1 MMHUB1	laminin beta-1 cha
26	87	5.4	290	2 G72858	AcOrf-70 protein -
27	87	5.4	501	2 S56163	tumor necrosis fac
28	87	5.4	652	2 T38704	glycyl tRNA synthe
29	87	5.4	902	2 T01127	curly leaf protein

30 86.5 5.4 4545 1 S25111
31 86 5.4 152 2 T18975
32 86 5.4 610 2 A35046
33 86 5.4 830 2 A30359
34 86 5.4 1034 2 JC5598
35 85.5 5.3 769 1 IJHULM
36 85.5 5.3 772 2 S32659
37 85.5 5.3 806 2 A46271
38 85.5 5.3 1599 2 T16210
39 85 5.3 398 2 A35281
40 85 5.3 2543 2 T31687
41 84.5 5.3 169 1 S18946
42 84.5 5.3 307 2 T09923
43 84.5 5.3 1639 1 MMFPB2
44 84 5.3 769 1 JC1121
45 84 5.3 788 2 I51530

. ALIGNMENTS

RESULT 1

T23152

hypothetical protein JC8.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23152

R:Lightning, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19699

A:Accession: T23152

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-429 <WIL>

A:Cross-references: EMBL:Z82274; PIDN:CA805228.1; GSPDB:GN00022; CESP:JC8.6b

A:Experimental source: clone JC8

C:Genetics:

A:Gene: CESP:JC8.6b

A:Map position: 4

A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Query Match 24.8%; Score 396; DB 2; Length 429;
Best Local Similarity 38.9%; Pred. No. 1.2e-25;
Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

Qy 98 YCDFSSGDFCNSCSC-----NNLRHELEFRFAIKACLDNPEAPQPKMKGRLGAKLR- 152

Db 182 YCDCFANGFECDNCCKDCHNNIEYDSORSKAIROSLERNPNAPKPKIGIARGGTIDIER 241

Qy 153 -HSKGCNCRSGCLKNKYCEYEAKEIMCSSICKIACKN-----YEES---PERKML 199

Db 242 LHQKCHCKKSGCLKNKYCEYEAKEVPCDCKCKGCQNTYRTYRKNSSGGAYSNVNL 301

Qy 200 MS-----TPHYMEPGDFESSHY---LSPAKFSGPPKLRKNRQAFSCISW-----EVV 243

Db 302 MSLTNASTATPD-SGPGSVTDEHGDDYEDMLLSHKPKVEMDPRRF---PWYMTDEVV 357

Qy 244 EATCACLAAQGEA-----EQEHCSPSLAEQMILLEEGRCLSQIL 283

Db 358 EATMCWVAQAEALNYEKVQTEDEKLNMEKLVLEFRGRCLEQMI 403

RESULT 2

B84585

hypothetical protein Atg20110 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: B84585

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,

euss, D.; Nierman, W.C.; Whizler, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84585

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-571 <STO>

A:Cross-references: GB:AE002093; NID:g4580462; PIDN:AAD24386.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g20110

A:Map position: 2

Query Match 22.7%; Score 363; DB 2; Length 571;

Best Local Similarity 32.4%; Pred. No. 9.5e-23;

Matches 81; Conservative 39; Mismatches 76; Indels 54; Gaps 8;

QY 36 SSFQSELPKPMPTLVGRLL-----PVPKALNLTQVD-NGALPSAVNGAAF 81

DB 46 SSMAASIPSPVTVRPIITTSQAPTVATPIPPPOSQIILHVRHPRPESPSMMPR 105

QY 82 PSGPALQGPVK-----ITLSGYDCDFSGDFCNCSGSC-----NNLRHELEPRKAIK 127

DB 106 PAGETRDGTQKKKQCKNCKHSRCLKLYCECFASGTCDGNCVNCFNVENEPARQAVE 165

QY 128 ACIDRNPEAFQPKMGKGRGLGAAK-----LHSGKCNCKRSGLKNYCEYEAKI 176

DB 166 STLERNPAPRPKTAASPHGGRDNREVDVVMRLARHNKCHCKKSKLKYCECFQANI 225

QY 177 MCSICKCIACKNVEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSGP 225

DB 226 LCSENCKCLCKNFESEVRQSLFHGHSNLMAYLQOANNAITGAVGSSGFAS-----APP 281

QY 226 PKLRKNQAF 235

DB 282 PKRRKGQEIF 291

RESULT 3

T08955

hypothetical protein F19B15.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999

C:Accession: T08955

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,

submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16519

A:Accession: T08955

A:Molecule type: DNA

A:Residues: 1-603 <BEV>

A:Cross-references: EMBL:AL078470; GSPDB:GN00062; ATSP:F19B15.30

A:Experimental source: cultivar Columbia; BAC clone F19B15

C:Genetics:

A:Gene: ATSP:F19B15.30

A:Map position: 4

A:Introns: 107/2; 203/3; 321/3; 356/2; 388/1; 522/1

Query Match

Best Local Similarity 22.3%; Score 357; DB 2; Length 603;

Matches 88; Conservative 38; Mismatches 66; Indels 58; Gaps 12;

QY 36 SSFQSELPKPMPTLVGRLLPVPKALNLTQVDNGALPSAVNGA----APPSPGALQGP- 90

DB 60 SSVFSTIRP-GMTAIGQVTVQVRPTLPMTATMSNPSPQSOIVNAPTRHPSPKARGPR 118

QY 91 PKI-----TLSGYDCDFSGDFCNCSGSC-----NNLRHELEPRKAIK 127

DB 119 PNVEGRDGTQKKKQCKNCKHSRCLKLYCECFASGTCDGNCVNCFNVDNEPARREAVE 178

QY 128 ACIDRNPEAFQPKMGK-----LGAALKL--RHSGKCNCKRSGLKNYCEYEAKI 176

DB 179 ATLERNPAPRPKTAASPHGGRDKREDIGEWVLLGHNKCHCKKSKLKYCECFQANI 238

QY 177 MCSICKCIACKNVEESPERKMLMSTPH-----YME-----PGDFESSHYL-SPAKFS 223

DB 239 LCSENCKCLCKNFESEVRQSLFHGHSNLMAYLQOANNAITGAVGSSGFAPSPA--- 295

QY 224 GPPKLRKNRQ 233

DB 296 --PKRRKGQE 303

RESULT 4

F71410

hypothetical protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

A:Variety: Columbia

C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998

C:Accession: F71410

R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D

avanagh, T.; Hempel, S.; Kottler, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B

Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen

erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;

C.; Chalwatzis, N.

A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t

A:Reference number: A71400; MUID:98121113; PMID:9461215

A:Accession: F71410

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-658 <BEV>

A:Cross-references: GB:Z97337; NID:g2244829; PID:e326824; PID:g2244834

C:Genetics:

A:Map position: 4COP9-4G3845

Query Match

Best Local Similarity 14.5%; Score 232; DB 2; Length 658;

Matches 51; Conservative 11; Mismatches 43; Indels 24; Gaps 5;

QY 82 PSGPALQ-----GPPKITLSGYDCDFSGDFC-NSCSC-----NNLRHELEPRKAIKACLD 131

DB 354 PVEPALQELNLSPKK---KSYCEFAAGVYCEPCSCIDCFNKPITHEDVVLATRKQIES 410

QY 132 RNPEAFQPKMGKGRGLGCA-----AKLRHSGKCNCKRSGLKNYCEYEAKIMCSS 180

DB 411 RNPLAFAPKVRNSDSVQETGDDASKTPASARHKRCNCKKSNCLAKKYCECYGGVGCSE 470

QY 181 IKCICACKN 189

DB 471 NCRCEGCKN 479

RESULT 5

T00209

MEGF8 protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00209

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like moti

A:Reference number: Z14126; MUID:98360089; PMID:9693030

A:Accession: T00209

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: mRNA

A:Residues: 1-1737 <NAK>

A:Cross-references: EMBL:AB011541; NID:g3449307; PIDN:BAA32469.1; PID:g3449308

A:Experimental source: brain; clone HG1392

C:Genetics:

A:Gene: MEGF8

A:Map position: 19q12

Query Match

Best Local Similarity 7.2%; Score 115; DB 2; Length 1737;

Matches 73; Conservative 27; Mismatches 124; Indels 92; Gaps 18;

QY 17 NCGAREKALHLLLPQYDDQSSFPQSELPKPRMTTLVGRLLPVPKALNLTQVDNGALPSAV 76

Db 711 NCNA-----WLPDITRAS-----VGPPMEESVAHAAGVSRILYISGGFGGVALGRL 759
 QY 77 NGAAPSPGA--LQGPPIKTLGYCD-CFS---SGD-----FCNSCSNNL---RHELER 122
 Db 760 LALTLPDPCRLSSPEACNQSGACTWCHGACLSGQAHRLCGGSGCSPMPSPRECR 819
 QY 123 FKAIAKACLDNRNPAFOPKMGKRLGAALKRHSKGC-NCKRSOCL-----KNYC-----169
 Db 820 LRTCSECLARHPTLQPGDGE-----ASTPRCKWCTNCPGACIGRNGSCTSENDCRINQ 874
 QY 170 -----ECYEAKI-----MCSSICKCIACKNYEESPE-RKMLMSTPHY-----205
 Db 875 REVFWAGNCESEACGAADCEQCTREKCMWTRQFKRTGETRILSVQPTDYDMTCFHSLL 934
 QY 206 -MEPGDFESSHYLSPAKFSGPPKLRNQAFCISWVEVATCACLLAOGAEAEQHCSP 264
 Db 935 NVSPMPVSS---PPLPCTTPCHLLPN-----CTS-----CLDSKGADGGWQHCVW 977
 QY 265 SLAEOQMLEEFGRCLS 280
 Db 978 SSSLQ-----QCLS 986

RESULT 6
 T21436
 hypothetical protein R06A4.7 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T21436; T23953
 R:Barlow, K.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19421
 A:Accession: T21436
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-775 <WIL>
 A:Cross-references: EMBL:281515; PIDN:CAB04199.1; GSPDB:GN000020; CESP:R06A4.7
 A:Experimental source: clone F26H11
 R:Barlow, K.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19822
 A:Accession: T23953
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-775 <WIL>
 A:Cross-references: EMBL:283120; PIDN:CAB05589.1; GSPDB:GN000020; CESP:R06A4.7
 A:Experimental source: clone R06A4
 C:Genetics:
 A:Gene: CESP:R06A4.7
 A:Map position: 2
 A:Introns: 223/1; 414/3; 464/2; 537/1; 601/2; 655/1; 679/1; 749/2

Query Match 6.3%; Score 100; DB 2; Length 775;
 Best Local Similarity 21.8%; Pred. No. 1.6;
 Matches 32; Conservative 14; Mismatches 41; Indels 60; Gaps 7;

QY 89 GPPKITLGYCDGCFSSGDFCNSCSNNLRHELERFAIKACLDNRNPAEFQPKMGKRLGA 148
 Db 537 GPCNATAEN-CACRENGVCSYCKCD-----INCSQRF-----569
 QY 149 AKLRHSGCNCKRSGCLKNYCEYEAIKMCS-----SICKC-----IACKNYEESPERK 197
 Db 570 -----GCNCAAGQCYTRACQCYRANWCNPNWTCNWKCDIDSNIKRNFGWT---R 619
 QY 198 MLMSTPHYMEPGDFESSHYLSLPAKFSG 224
 Db 620 MI-----QKRTYCGPSKIAG 634

RESULT 7
 T27827
 hypothetical protein ZK287.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T27827
 R:McMurray, A.
 submitted to the EMBL Data Library, April 1996
 A:Reference number: Z20425
 A:Accession: T27827
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-497 <WIL>
 A:Cross-references: EMBL:270757; PIDN:CAA94805.1; GSPDB:GN000023; CESP:ZK287.1
 A:Experimental source: clone ZK287
 C:Genetics:
 A:Gene: CESP:ZK287.1
 A:Map position: 5
 A:Introns: 20/1; 44/2; 99/3; 339/3; 443/2

Query Match 6.1%; Score 97; DB 2; Length 497;
 Best Local Similarity 26.6%; Pred. No. 1.8;
 Matches 37; Conservative 15; Mismatches 51; Indels 36; Gaps 10;

QY 82 PSGALQGPPIKTLGYCDGCFSSGDFCNS-CSCNNLRHELERFAIKACLDNRNPAEFQPK 140
 Db 63 PAKPAKNACAK---SERCDLRIKCGCKLCAKQI-----C--KNKEA--PK 103
 QY 141 MGKGRGLGAALK-RHSKGCNCK--RSGCLKNYCEYEAIKMCSICKICIA-CKNYEESPER 196
 Db 104 K-----LAKVAKPTSGCCAKGKKQCVKKECACRTVYGCASCKCGGCTNGASK--- 154
 QY 197 KMLMSTPHYMEPGDFESSH 215
 Db 155 ---FSVPKHVQNCFLHKK 170

RESULT 8

MMFFBI

laminin beta-1 chain precursor - fruit fly (Drosophila melanogaster)

N:Alternate names: laminin chain B1

C:Species: Drosophila melanogaster

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001

C:Accession: A28783; SI4462; B28783

R:Montell, D.J.; Goodman, C.S.

Cell 53, 463-473, 1988

A:Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals

A:Reference number: A28783; MUID:88210471; PMID:3365769

A:Accession: A28783

A:Molecule type: mRNA

A:Residues: 1-1790 <MON1>

A:Cross-references: EMBL:M19525

R:Montell, D.J.; Goodman, C.S.

submitted to the EMBL Data Library, June 1988

A:Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain r

A:Reference number: SI4462

A:Accession: SI4462

A:Molecule type: mRNA

A:Residues: 1-667,'L',669-725,'VT',728-947,950-1790 <MON2>

A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802

C:Genetics:

A:Gene: lambl

A:Cross-references: FlyBase:FBgn0002527

A:Map position: 2L 28D

C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin

C:Function:

A:Description: interact with cells and with other basement membrane proteins to promo

C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology

C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>

F:27-288/Domain: VI <DOM6>

F:289-561/Domain: V <DOM5>

F:290-354/Domain: laminin-type EGF-like homology <LE01>

F:357-417/Domain: laminin-type EGF-like homology <LE02>

F:420-477/Domain: laminin-type EGF-like homology <LE03>

F:480-528/Domain: laminin-type EGF-like homology <LE04>

F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>

F:562-789/Domain: IV <DOM4>

F:643-645/Region: cell attachment (R-G-D) motif

F:790-1189/Domain: III <DOM3>

F:791-836/Domain: laminin-type EGF-like homology <LE06>

F:839-882/Domain: laminin-type EGF-like homology <LE07>

F:885-932/Domain: laminin-type EGF-like homology <LE08>

F:935-990/Domain: laminin-type EGF-like homology <LE09>

F:968-972/Region: cell adhesion #status predicted

F:993-1042/Domain: laminin-type EGF-like homology <LE10>

F:1045-1093/Domain: laminin-type EGF-like homology <LE11>

F:1096-1141/Domain: laminin-type EGF-like homology <LE12>

F:1144-1188/Domain: laminin-type EGF-like homology <LE13>

F:1190-1407/Domain: II <DOM2>

F:1408-1434/Domain: alpha <ALP>

F:1435-1790/Domain: I <DOM1>

F:51-56/Disulfide bonds: #status predicted

F:140-203,234,489,593,1053,1249,1303,1332,1343,1475,1495,1517,1583,1646,1705/Binding site

F:1191,1194,1788/Disulfide bonds: interchain #status predicted

Query Match 6.0%; Score 96.5; DB 1; Length 1790;

Best Local Similarity 21.0%; Pred. No. 7.2;

Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;

QY 94 TLSGYCDGCFSSGDF-----CNSCSC-----NNLRHELEFKAIFKACLDNRNPEAFQPKM 141

Db 1021 TTGDHCELCKDGFQDALQNCQCEDFLTGNTTIAHCDRFTGQCPCPCL-----PNV 1072

QY 142 GKRLGAALRHSHK-----GCNCRSGCLKNYCEYAKIMCSTICKCIACK-----N 189

Db 1073 QGVRCDOCAENHWHKIASGEGCESCNDPIGALHEQCNSYTGCCCKPFGGRACNQCQAH 1132

QY 190 YEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPKLRKNQAFSCISWEVVEATCAC 249

Db 1133 YWGNPNEKQPCCECDQFGAADFQC-----DRETGNVCVCHGEGGYKCN 1175

QY 250 LLAOGERAEQEHCSF 264

Db 1176 ECARGYIGQFPCHSP 1190

RESULT 9

S55554

male-specific lethal 2 protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000

C:Accession: S55554

R:Zhou, S.; Yang, Y.; Scott, M.J.; Pannuti, A.; Fehr, K.C.; Eisen, A.; Koonin, E.V.; Foul

EMBO J. 14, 2884-2895, 1995

A:Title: Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex

A:Reference number: S55554; MUID:95317307; PMID:7796814

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-769 <ZHO>

A:Cross-references: EMBL:248443; NID:9872110; PIDN:CAA88358.1; PID:9872111

C:Genetics:

A:Gene: FlyBase:msl-2

A:Cross-references: FlyBase:FBgn0005616

A:Introns: 641/1

Query Match 6.0%; Score 96; DB 2; Length 769;

Best Local Similarity 25.7%; Pred. No. 3.3;

Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;

QY 121 ERFKAIKAC-----LDRNPEAFQPKMGRGLGAALRHSGK-----CNCKRSRG 163

Db 471 KRTRLKASQAQAEKLEVPSEVTKTVQSGK-GA--LRRIRGKDKKEKVKPKPKRCRGISG 527

QY 164 -----CLKNYCEYAKIMCSTICKCIACKN-----YEES-----PER 196

Db 528 SSNTLTTCRNSRCFCYKSYNSCAG-CHVCCKNPKHKEDYVESDEDDLEDFEMPKDYPEP 586

QY 197 KMLMSTPHYMEPGDFESSHYLSPAKFSGPPKL 228

Db 587 MTQSEEPVVAEPROEENS--MAPDSSAPISL 616

RESULT 10

T10053

laminin alpha 5 chain - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000

C:Accession: T10053

R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.

submitted to the EMBL data Library, November 1997

A:Reference number: Z16923

A:Accession: T10053

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3635 <MIN>

A:Cross-references: EMBL:U37501; NID:g2599231; PID:g2599232

C:Genetics:

A:Gene: Lama5

C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like h

C:Keywords: basement membrane; cell binding; extracellular matrix

F:1888-1939/Domain: laminin-type EGF-like homology <LEG>

F:1942-1970/Domain: EGF homology <EGF>

Query Match 6.0%; Score 96; DB 2; Length 3635;

Best Local Similarity 18.0%; Pred. No. 17;

Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;

QY 14 CIDNCGARELKALHLLPQYDQCSFPOSELPKPMITLVGLRLLPVPKLNILITVDNGLP 73

Db 117 CLERFGPRTLER---ITQDDVI-----CTTEYSRIVPL-ENGEIVVSLVNGR-P 161

QY 74 SAVNGAARF-----SGPALQGP-----PKITLSGY 98

Db 162 GALNFYSYPLLRDFTKATNIRLPLRNTLLGLMLGRALRDPVTVRRYYYSIKDISIGGR 221

QY 99 CDCFSFGDFC-----NSC-----SC 113

Db 222 CVCHGHADVCDAKDPLDPLRLOCAQHNTCGGSCDRCPGPNQPKPATITSDANECQSC 281

QY 114 NLRH-----ELERFKAIKACLDNRNPEAFQPKMGKGRGLGAAL---RHSKGCNCKRS 162

Db 282 NCHGHAYDYDPEVDR-----RNASQNDNVYQG--GGVCLDCQHHHTTGINGER-- 329

QY 163 GCLKNY-----CEYEAKI---MCSSTI---CK-----CIAC-KNY 190

Db 330 -CLPGFFRAPDQPLDSPHVCRCDC-ESDFTDGTCEDLTGRCYCRPNFTGELCAACAGY 387

QY 191 EESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPKLRKNQAFSCISWEVVEATCAC 250

Db 388 TDPHCYPLPSFPHN-----DTREQVLPA-----GGIVNCDNAA 422

QY 251 LAQGEAEQEQ-----HC-----SPSLAEQMLLEFGR 277

Db 423 GTQGNACRKDPRLGRVCVKPNFRGAHCELCAFGHPSCHPCQCSGPGVANSIDPESGQ 482

QY 278 CL 279

Db 483 CM 484

RESULT 11

S61917

protein kinase C (EC 2.7.1.1) PKCA - Aspergillus niger

C:Species: Aspergillus niger

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S61917

R:Morawetz, R.; Lendenfeld, T.; Mischak, H.; Muehlbauer, M.; Gruber, F.; Goodnight, J

Mol. Gen. Genet. 250, 17-28, 1996

A:Title: Cloning and characterisation of genes (pkcl and pkca) encoding protein kinas

A:Reference number: S61917; MUID:96158841; PMID:8569684

A:Accession: S61917
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1096 <NOR>
A:Cross-references: EMBL:U10549; NID:g507899; PIDN:AAA97433.1; PID:g507900
C:Genetics:
A:Gene: pkcA
C:Introns: 145/1; 234/1; 428/1; 560/1; 698/1; 874/2; 941/3; 1077/1
C:Superfamily: yeast protein kinase C; protein kinase C zinc-binding repeat homology; protein kinase C
C:Keywords: ATP; duplication; phospholipid binding; phosphotransferase; serine/threonine kinase
F:460-507/Domain: protein kinase C zinc-binding repeat homology <K21>
F:528-577/Domain: protein kinase C zinc-binding repeat homology <K22>
F:769-1030/Domain: protein kinase homology <KIN>
F:777-785/Region: protein kinase ATP-binding motif

Query Match 5.9%; Score 95; DB 1; Length 1096;
Best Local Similarity 22.6%; Pred. No. 5.8;
Matches 63; Conservative 21; Mismatches 91; Indels 104; Gaps 15;

Qy	66	QVONGALPSAVN-----GAFFPSGALQPPKKTLSGYCDCFSSGDFCNSCCSNLURHEL	120
Db	359	KMEHGGAAPRPDTAGGAGSAHSPGCG--GFGAAPSGQYDGSGAGAPAGNQVMTDSW-FAL	415
Qy	121	ERFKAI-----KACLDNRPEAFQPKMGKGRUGAAKLRL-----HSGK-----C	157
Db	416	EPVGRIVLSMSFAKOLKDRPF-----DIGNRQAGAVRQKKEEVEHQGHKFVTOQFYNTM	471
Qy	158	NCRSGCLKNYC---ECYEAKIMCSTIC-----KCIACKNYPESPCKMLMSTPHYMEP	208
Db	472	RCALCGDFLYAAGMOCADCKYTCCHKKYPKVYTKGISKANYETDP-----	517
Qy	209	GDFESSHYLSPAKFSPPKLRKRNQAFSCISWEVVEATCACLIAQGEAEAEQHCSPSLAE	268
Db	518	-DEEKINHRIPHREF-----FSNTS-----ANWCC-----HCG-----	545
Qy	269	QMILEEFGR-----CLSQILHIEFKSKGLKIE	295
Db	546	--VLLPFGRNAKRCETGCTHSHGCTHLVPDFCGMSME	582

RESULT 12

T08852

lustrin A - California red abalone

C:Species: *Halioctis rufescens* (California red abalone)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: T08852

R:Shen, X.; Belcher, A.M.; Hansma, P.K.; Stucky, G.D.; Morse, D.E.

J. Biol. Chem. 272, 32472-32481, 1997

A:Title: Molecular cloning and characterization of lustrin A, a matrix protein from shell

A:Reference number: Z16496; MUID:98070424; PMID:9405458

A:Accession: T08852

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1428 <RSE>

A:Cross-references: EMBL:AF023459; NID:g2723361; PIDN:AA895154.1; PID:g2723362

A:Experimental source: tissue type mantle (shell and pearl nacre); cell type pallial

C:Superfamily: antileukoproteinas repeat homology

C:Keywords: extracellular matrix; extracellular protein

F:1382-1426/Domain: antileukoproteinas repeat homology <ALP>

Query Match 5.9%; Score 95; DB 2; Length 1428;
Best Local Similarity 22.8%; Pred. NO. 7.7;
Matches 56; Conservative 21; Mismatches 81; Indels 88; Gaps 14;

```

QY 74 SAVNGAAPPSPAL- ---QGPPKTLSCYDCFSS-----GFCN-----S 110
    | | | | |
Db 409 SCVGAACPRNTCAAPSAPV- ---C-CYTSGPPRPPSPPTGDCPQVNVN 462
    | | | | |
QY 111 CSCNNLR- ---HELERFAIKACLDORNEAFQP ---KMGKRGRLGAAKLRSHKGCNCKRS 162
    | | | | |
Db 463 CTAGTCRLVEDCRIRGCPAVPKCIDROPPLPPDPVCPGTPVLGA ---DLKQIYCGRG 518
    | | | | |
QY 163 G--GLKN-YCSCYIFA-----KIMGS-SICKCIA 186

```

```

      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db    519 GRRCPWNTYCVIHPADRYAVCCFGSGPASAIAPTSAPGVPDPCEPGVNVNACTIGVCRLVA 578
Qy    187 CKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLKRNRQAFCISWEVVVEAT 246
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db    579 NCDTWPCPARTCVD--HSPEP-----SLNCTIGDPALNGKLEEFSCVGGRICPLN 627
Qy    247 CACILLA 252
      ||| |
Db    628 TACLAA 633

RESULT 13
T26998
hypothetical protein Y48B6A.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26998
R;Wall, M.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z20297
A;Accession: T26998
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-716 <WIL>
A;Cross-references: EMBL:AL110490; NID:e1542263; PIDN:CAB54442.1; CESP:Y48B6A
A;Experimental source: clone Y48B6A
C;Genetics:
A;Gene: CESP:Y48B6A.6
A;Introns: 38/3; 196/3; 437/2; 460/3; 518/1; 617/3; 673/3

```

Query Match 5.98; Score 94; DB 2; Length 716;

Best Local Similarity	25.08	Pred. No. 4.6;	95;	Indels	46;	Gaps	10;
Matches	58;	Conservative	33;	Mismatches			
Qy	22	ELKALHLLPQYDDQ-----SSFPOSELPKPMWTLVGRLL--LPVPAKLNLITQVD-	68				
Db	378	EVKALEAQESDDRPSAPTPTRDSLSLPPPPPKPTPLAIRRAGIPTPQLLEMIHQEDC	437				
Qy	69	--NCALPSAVNGAAFFPSGALQGGPKPITLSGYDCDFSCNCSNNLRHLEFRKAI	126				
Db	438	SIRFSSPTSVSHGSRQSPAPVKKPSVTVS-----PLGLLCDP---NL--STIEKPEEM	485				
Qy	127	KACLDNRPEARQPKMGKRGRLCAAKLRHSHGCNCKRSGCLKNICYCEYBAKIMCSSICKCIA	186				
Db	486	KTETKPVETAPAPDEAEALNDALDRNNK---INEATCL-----TKIFPSLVSKYFQ	534				
Qy	187	CKNYEESPERKMLMSTPHYMPGGDFESSHYLSPAKFSFGPKPLKRNQRAFSCI	238				
Db	535	NCSPDFDLNKKRSGNGVPLKMNK---KISLYAEFSFSS-----RKQIOYFSGI	578				

RESULT 14

A71928
cag island protein - *Helicobacter pylori* (strain J99)
C:Species: *Helicobacter pylori*
A:Variety: strain J99
C:date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: A71928
R:Alm, R.A.; Ling, L.S.-L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.;
Ives, C.; Gibson, R.; Merberg, D.; Malls, S.D.; Jiang, Q.; Taylor, D.E.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human
A:Reference number: A71800; PMID:99120557; PMID:9923682

A;Residues: 1-1819 <ARN>

A; Cross-references: GB:AE001481; GB:AE001439; MID:g4155005; PIDN:AAD06047.1; PID:g4155005
A; Experimental source: strain J99
C; Genetics:
A; Gene: orf13/14

Query Match 5.9%; Score 94; DB 2; Length 1819;
 Best Local Similarity 21.4%; Pred. No. 12;
 Matches 63; Conservative 39; Mismatches 113; Indels 80; Gaps 12;

QY 22 ELKALHLLPOYDDQSFPOSELKP-----MTLVCRLLPVPAK-----LNLIQVDNG 70
 DB 568 DILAKESLAKYDCASQAKTEAEKCEKLLTPPEAKLLLEAEKESVKAYLDCVSKAQTE 627
 QY 71 ALPSAVNGAFAFPGPALQGPPIITLGGYDCDFSSGDFCNSCNCNNRHELEKFAKACKL 130
 DB 628 AEKKECEKLLTPPAKKLLEAKSVAYLDCVSKA-----KNAERKECEKLL- 675
 QY 131 DRNPEAFQPMQGRIGAAKLHRSKQNC-----KSGCLKNKYCEYAKIMCSCSIC 182
 DB 676 --TPE-----AKLLENQALDCLKNAKTDEERKECKLDPKDLQKVLAKESV 721
 QY 183 K-----CIA-CKNYEESPERKMLMSTHYMEPCDFESSHYLSYSPAKFSPPKLRKNQAFSC 237
 DB 722 RYVLDVSKAKNAEAKKECKLL-TP-----EARKLLEAKKS----- 758
 QY 238 ISWEVVEATCACLLAQGEAEQHCSPSLAEQ--MILEEFGRCLSOILHIEFKSK 290
 DB 759 -----VKAYKDCVLRARNEKEKECEKLLTPPEAKLLEESKSKSVKAYLDCVSKAK 808

RESULT 15
 S68960
 laminin alpha-4 chain precursor - human
 N:Alternate names: laminin Ah
 C:Species: Homo sapiens (man)
 C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 16-Jun-2000
 C:Accession: S68960; S65926; S49149; S40150; I53516
 R:Richards, A.; Al-Imara, L.; Pope, F.M.
 Eur. J. Biochem. 238, 813-821, 1996
 A:Title: The complete cDNA sequence of laminin alpha-4 and its relationship to the other
 A:Reference number: S68960; MUID:96300249; PMID:8706685
 A:Accession: S68960
 A:Molecule type: mRNA
 A:Residues: 1-1816 <RIC>
 A:Cross-references: EMBL:X91171; NID:g1212962; PIDN:CAA62596.1; PID:g1212963
 A:Experimental source: tissue type heart
 R:Ilivanainen, A.; Sainio, K.; Sariola, H.; Tryggvason, K.
 FEBS Lett. 365, 183-188, 1995
 A:Title: Primary structure and expression of a novel human laminin alpha-4 chain.
 A:Reference number: I53516; MUID:95300971; PMID:7781776
 A:Accession: S65926
 A:Molecule type: mRNA
 A:Residues: 1-142, 'P', 144-177, 'F', 179-490, 'Y', 492-1056, 'P', 1058-1816 <IIIV>
 A:Cross-references: EMBL:S78569; NID:g1042081; PIDN:AB34635.1; PID:g1042082
 R:Richards, A.J.; Al-Imara, L.; Carter, N.; Leversha, M.; Lloyd, J.C.; Pope, F.M.
 submitted to the EMBL Data Library, December 1993
 A:Description: Localisation of the gene (LAMA4) to chromosome 6q21 and isolation of a pa
 A:Reference number: S49149
 A:Accession: S49149
 A:Molecule type: mRNA
 A:Residues: 236-1816 <RI2>
 A:Cross-references: EMBL:X76939; NID:g509805; PIDN:CAA54258.1; PID:g509806
 R:Richards, A.J.; Al-Imara, L.; Carter, N.; Lloyd, J.C.; Pope, F.M.
 submitted to the EMBL Data Library, February 1993
 A:Description: Isolation of a partial cDNA encoding a protein homologous to laminin A.
 A:Reference number: S40150
 A:Accession: S40150
 A:Molecule type: mRNA
 A:Residues: 1403-1541, 'S', 1543-1816 <RI3>
 A:Cross-references: EMBL:X70904; NID:g437804; PIDN:CAA50261.1; PID:g437805
 C:Genetics:
 A:Gene: GDB:LAMA4; LAMA3
 A:Cross-references: GDB:203904; OMIM:600133
 A:Map position: 6q21-6q21
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
 C:Function:
 A:Description: Interact with cells and with other basement membrane proteins to promote
 C:Superfamily: laminin alpha-4 chain; laminin G repeat homology; laminin-type EGF-like H

C:Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glyco
 F:1-24/Domain: signal sequence #status: predicted <SIG>
 F:25-1816/Product: laminin alpha-4 chain #status: predicted <MAT>
 F:82-129/Domain: laminin-type EGF-like homology <LE1>
 F:132-184/Domain: laminin-type EGF-like homology <LE2>
 F:187-238/Domain: laminin-type EGF-like homology <LE3>
 F:241-265/Domain: laminin-type EGF-like homology #status: atypical <LE4>
 F:717-719/Region: cell attachment (R-G-D) motif
 F:862-1031/Domain: laminin G repeat homology <LG1>
 F:1068-1223/Domain: laminin G repeat homology <LG2>
 F:1252-1398/Domain: laminin G repeat homology <LG3>
 F:1488-1636/Domain: laminin G repeat homology <LG4>
 F:1665-1816/Domain: laminin G repeat homology <LG5>
 F:104,215,308,458,524,550,571,574,631,639,735,751,754,780,803,1086,1281,1359,1411/Bin
 F:266,269/Disulfide bonds: Interchain #status: predicted

Query Match 5.8%; Score 93; DB 1; Length 1816;
 Best Local Similarity 23.8%; Pred. No. 14;
 Matches 74; Conservative 32; Mismatches 143; Indels 62; Gaps 18;

QY 2 VICQLKGAQMLCIDNCG-----ARELKALH---LLPOY---DDQSFPOSELKPMTTLV 51
 DB 80 VPCDCNGSNE-CLDGSYCVHCQRNTTGEHCEKCLDYGIDSGIRGAPQFCQPCP----- 133
 QY 52 GRLLPVPKLNLTQV--DNGALPSAVN-GAAPPS-----GPAQQPPKITLSGYDCFCSS 104
 DB 134 ---CPLPHLANFAESCVRKNGAVRCICNENYAGPNCERCAPGYGNPLTI----- 180
 QY 105 GDFCNSCSN-----NLRIE--LEREFAKACLDNRPEAFQ-PKMGKGRIGAAKL-RHSK 155
 DB 181 GSTCKKDCSGNSDPLIFEDCDDEVGTGQCRNCL-RNTTGFKCBRCAPGYIGDARIANCA 239
 QY 156 GCNCKRSGCLKNKYCEYAKIMCSCSIC-RCI-----ACKNYEESPERKMLMST--P 203
 DB 240 VNCGGGCPDVSVTGCECLEBGFEPPTGCDKCVWDLTDLLRLAALSIEGKSGVLSVSSGA 299
 QY 204 HYMEPGDFESSHYLSYSPAKFSPPKLRKNQAFSCISWEVVEATCACLLAQGEAEQHC 263
 DB 300 AHRHVNEINATIIYLLKTLSE-----RENOYALRKIQINNAENTMKSLLSDVEELVEKENQ 355
 QY 264 PSLAEQWILEE 274
 DB 356 ASRKGLVQKE 366

Search completed: April 21, 2003, 11:50:39
 Job time : 21.3956 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:43:43 ; Search time 27.8114 Seconds
(without alignments)
2185.571 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGGAGMLCIDNCGA.....GRCLSOILHTEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	100.0	475	11	O8VIE0 mus musculus
2	1589	99.4	475	11	O8VIE1 mus musculus
3	504	31.5	277	11	O9D571 mus musculus
4	435	27.2	950	5	O9V608 Q9V608 drosophila
5	396	24.8	429	5	O62295 caenorhabdi
6	396	24.8	435	5	O950D7 Q950d7 caenorhabdi
7	363	22.7	571	10	O9SL70 Q9SL70 arabisopsis
8	357	22.3	603	10	O9SZD1 Q9SZD1 arabisopsis
9	253.5	15.9	356	10	O9LW71 Q9LW71 arabisopsis
10	251.5	15.7	601	10	O9CAV1 Q9CAV1 arabisopsis
11	232.5	14.5	896	10	O9ZS22 Q9ZS22 glycine max
12	232	14.5	658	10	O23333 Q23333 arabisopsis
13	223.5	14.0	609	10	O9M679 Q9M679 arabisopsis
14	222	13.9	526	10	O94A12 Q94A12 arabisopsis
15	222	13.9	695	10	O9LE32 Q9LE32 arabisopsis
16	222	13.9	695	10	O9LUI3 Q9LUI3 arabisopsis

17	186	11.6	243	5	O9VMQ3 Q9VMQ3 drosophila
18	180.5	11.3	593	10	O9LUI5 Q9LUI5 arabisopsis
19	159	9.9	553	10	O94DS2 O94DS2 oryza sativ
20	115	7.2	1737	4	O75097 O75097 homo sapien
21	101	6.3	769	11	O8VHL6 O8VHL6 sigmodon hi
22	100	6.3	773	5	O17514 O17514 caenorhabdi
23	100	6.3	773	5	O62335 O62335 caenorhabdi
24	99.5	6.2	1704	5	O94446 O94446 chironomus
25	99	6.2	2931	5	O9W2C6 O9W2C6 drosophila
26	97.5	6.1	772	13	O9PU04 O9PU04 ictalurus p
27	97	6.1	497	5	O23460 O23460 caenorhabdi
28	97	6.1	984	5	O9Y1P7 O9Y1P7 cryptospori
29	97	6.1	1106	5	O9GU55 O9GU55 drosophila
30	97	6.1	1987	5	O9V841 O9V841 drosophila
31	96.5	6.0	1026	5	O8SWY0 O8SWY0 drosophila
32	96.5	6.0	4547	5	O9W343 O9W343 drosophila
33	96	6.0	773	5	O9VQR1 O9VQR1 drosophila
34	95.5	6.0	448	5	O9NKE0 O9NKE0 drosophila
35	95.5	6.0	1035	5	O9VJ08 O9VJ08 drosophila
36	95	5.9	1428	5	O44341 O44341 haliotis ru
37	94.5	5.9	631	15	O9IEZ0 O9IEZ0 caprine art
38	94	5.9	716	5	O9U2A6 O9U2A6 caenorhabdi
39	94	5.9	1819	16	O9ZLV0 O9ZLV0 helicobacte
40	93.5	5.8	420	5	P91776 P91776 pacifastacu
41	93.5	5.8	1981	5	O9VVK7 O9VVK7 drosophila
42	93	5.8	463	11	O8ROC7 O8ROC7 mus musculu
43	93	5.8	761	5	O76733 O76733 drosophila
44	93	5.8	4123	4	O75851 O75851 homo sapien
45	92.5	5.8	489	11	O9D211 O9D211 mus musculu

ALIGNMENTS

RESULT 1

O8VIE0 ID O8VIE0 PRELIMINARY; PRT; 475 AA.
AC O8VIE0;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Tesmin (Hypothetical 50.6 kDa protein).
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Sutoh S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent protein; a possible multifunctional protein with dynamic changes of RT localization throughout spermatogenesis.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057423; BAB64935.1; -;
DR EMBL; BC024377; AAH24377.1; -;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Hypothetical protein.
SQ SEQUENCE 475 AA; 50615 MW; 4373ED1188DCF79F CRC64;

Query Match 100.0%; Score 1599; DB 11; Length 475;
Best Local Similarity 100.0%; Pred. No. 6.2e-161;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MVICQLKGGAGMLCIDNCGARELKALHLLPQYDDSSFPQSELPKPMVTTLVGRLLPVPK 60
Db 181 MVICQLKGGAGMLCIDNCGARELKALHLLPQYDDSSFPQSELPKPMVTTLVGRLLPVPK 240

```

QY 61 LNLITVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGCFSSGDFCNSCSCNNLRHEL 120
DB 241 LNLITVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGCFSSGDFCNSCSCNNLRHEL 300
QY 121 ERFKAIKACLDNRNPEAFQPKMGKRLGAALRHSKGCNCRSGCLKNKYCEYEAKIMCSS 180
DB 301 ERFKAIKACLDNRNPEAFQPKMGKRLGAALRHSKGCNCRSGCLKNKYCEYEAKIMCSS 360
QY 181 IKCIIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 240
DB 361 IKCIIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 420
QY 241 EVVEATCACLQAQEEAEQHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 295
DB 421 EVVEATCACLQAQEEAEQHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 475

RESULT 2
Q8VIE1
ID Q8VIE1 PRELIMINARY; PRT; 475 AA.
AC Q8VIE1;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE tesmin.
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RT "Structure of the tesmin gene encoding a testis-specific persistent protein; a possible multifunctional protein with dynamic changes of localization throughout spermatogenesis."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057422; BAB64934.1;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Query Match 99.4%; Score 1589; DB 11; Length 475;
Best Local Similarity 99.3%; Pred. No. 7.1e-160;
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MVICQLKGGQAQMLCIDNCGARELKALHLLPYDDQDSFFQSELPKPMPTLVGRLLVPAPK 60
DB 181 MVICQLKGGQAQMLCIDNCGARELKALHLLPYDDQDSFFQSELPKPMPTLVGRLLVPAPK 240
QY 61 LNLITVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGCFSSGDFCNSCSCNNLRHEL 120
DB 241 LNLITVDNGALPSAVNGAAPPSPALQGGPKITLSGYCDGCFSSGDFCNSCSCNNLRHEL 300
QY 121 ERFKAIKACLDNRNPEAFQPKMGKRLGAALRHSKGCNCRSGCLKNKYCEYEAKIMCSS 180
DB 301 ERFKAIKACLDNRNPEAFQPKMGKRLGAALRHSKGCNCRSGCLKNKYCEYEAKIMCSS 360
QY 181 IKCIIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 240
DB 361 IKCIIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPGPKLRNROAFSCISW 420
QY 241 EVVEATCACLQAQEEAEQHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 295
DB 421 EVVEATCACLQAQEEAEQHCSPSLAEQMILEEFGRCLSQLIHIEFKSKGLKIE 475

RESULT 3
Q9D571
ID Q9D571 PRELIMINARY; PRT; 277 AA.
AC Q9D571;

```

```

DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE adult male testis cDNA, RIKEN full-length enriched library,
DE clone:4930509C02, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015732; BAB29949.1;
SQ SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;

Query Match 31.5%; Score 504; DB 11; Length 277;
Best Local Similarity 100.0%; Pred. No. 4.2e-45;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVICQLKGGQAQMLCIDNCGARELKALHLLPYDDQDSFFQSELPKPMPTLVGRLLVPAPK 60
DB 181 MVICQLKGGQAQMLCIDNCGARELKALHLLPYDDQDSFFQSELPKPMPTLVGRLLVPAPK 240
QY 61 LNLITVDNGALPSAVNGAAPPSPALQGGPKITLSG 97
DB 241 LNLITVDNGALPSAVNGAAPPSPALQGGPKITLSG 277

RESULT 4
Q9V608
ID Q9V608 PRELIMINARY; PRT; 950 AA.
AC Q9V608;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG6061 protein.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

```

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Boltshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Dahle C., Dahle C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.C., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003818; AAF58365.1;
 DR FlyBase: FBgn0033846; CG6061.
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 SQ SEQUENCE 950 AA; 100021 MW; 627C3EA6B44A0A30 CRC64;

Query Match 27.2%; Score 435; DB 5; Length 950;
 Best Local Similarity 38.4%; Pred. No. 4e-37;
 Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

Qy 31 QYDQSSPQSELKPKMTTLVGRLLPVPKLNLTQVDNGALPVAAGAPFPGPALQGP 90
 Db 681 QLPTEQSTPIKVEKPLTLPQVKNVPAK--PLFEVLKPPATAAAGAVDPLGWTSSR 738
 Qy 91 PKITLSG-----YDCFSFGDFCNSCSC-----NNLRHELRFKAICLDRNPEAFQPK 140
 Db 739 KHCNCKSQCLKLYDCDFANGFCQDCCTCKDCFNLDYEVERAIRSCLDRNPSAFKPK 798
 Qy 141 MCKGRLGAALRHSGCKRGCKNYCEYEAQKIMCSICKCTACKNYEESPCKMLM 200
 Db 799 ITAPNSGDMRL-HNKGCKNSGCKNYCEYEAQKIMCSICKCTACKNYEESPCKMLM 852
 Qy 201 STPHYMEPGDFESSHYLSPAKFSGPKLRNQRQ-----AFSCISWEVVEATCACLAAQGE 255
 Db 853 -----VDMDSLGLAGVEQKKDK-AKNQLNENRANRYFTDDVIEATIMCHISRV 903
 Qy 256 EAEQHCSPSLAEQMIIEFGRCLSQIL 283
 Db 904 MHEKQNAVEDMERVEEMGESLTQII 931

RESULT 5
 ID O62295 PRELIMINARY; PRT; 429 AA.
 AC O62295;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE JCB.6b protein.
 GN JCB.6b.
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lightning J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z82274; CAB05228.1;
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 SQ SEQUENCE 429 AA; 48365 MW; 10601B22681C06D2 CRC64;

Query Match 24.8%; Score 396; DB 5; Length 429;
 Best Local Similarity 38.9%; Pred. No. 2e-33;
 Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

Qy 98 YDCFSFGDFCNSCSC-----NNLRHELRFKAICLDRNPEAFQPKMGRLGAALR- 152
 Db 182 YDCDFANGFCRDCNCKDCHNNIEYDSQSKAIRGSLERNPNAPKPKIGIARGGITDIER 241
 Qy 153 -HSKGCNCKRGCKNYCEYEAQKIMCSICKCTACKN-----YEES-----PERKML 199
 Db 242 LHQCHCKCKGCKNYCEYEAQKIMCSICKCTACKN-----YEES-----PERKML 301
 Qy 200 MS-----TPHYMEPGDFESSHY-----LSPAKFSGPKLRNQAESCISW-----EVV 243
 Db 302 MSLTNASSTATPD-SGPGSVVTDEHGDDYEDMLLSHKPKVEMDPRP---PWYMTDEVV 357
 Qy 244 EATCACLAAQGEA-----EQEHCSPSLAEQMIIEFGRCLSQIL 283
 Db 358 EAATMCWVAQAEALNYEKVQTEDEKLNMEKLVLEFGRCLEQMI 403

RESULT 6
 ID Q95QD7 PRELIMINARY; PRT; 435 AA.
 AC Q95QD7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE JCB.6a protein.
 GN JCB.6a.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lightning J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z82274; CAB05229.1;
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 SQ SEQUENCE 435 AA; 49109 MW; 7629A5C749D25A5D CRC64;

Query Match 24.8%; Score 396; DB 5; Length 435;
 Best Local Similarity 38.9%; Pred. No. 2.1e-33;

RA Pihaski-Maunsbach K., Marcker K.A., Jensen E.O.;
 RT "Cppl", a novel type DNA-binding protein involved in the expression of
 RL a soybean leghemoglobin c3 gene."
 DR Proc. Natl. Acad. Sci. U.S.A. 97:8163-8168(2000).
 DR EMBL: AJ010165; CA09028.1; -
 DR InterPro: IPR005172; CXG.
 DR Pfam: PF03638; CXG; 2.
 SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;

Query Match 14.5%; Score 232.5; DB 10; Length 896;
 Best Local Similarity 38.3%; Pred. No. 1.1e-15;
 Matches 44; Conservative 16; Mismatches 38; Indels 17; Gaps 3;
 YQ 98 YDCFSGGDFC-NSCSCNNLRHLERFKAI-----KACLDNRNPEAFQPKM----- 141
 DB 487 YDCFAAGTCTDPCACOGCLNRPVEVETVETKQIESRNPFAFKIVQPTTIDISSHM 546
 YQ 142 -GGRGLGAALRKSGCKNCRSGCLNKYCEYAKIMCSCICKIACKNVESPE 195
 DB 547 DDENLTPSSARHKGCKNCRSKLKYCEYQANVGCSSGCRGCKNVHGKKE 601

RESULT 12
 O23333 PRELIMINARY; PRT; 658 AA.
 AC O23333;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 72.1 kDa protein.
 GN AT4G14770.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98121113; PubMed=9461215;
 RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
 RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
 RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
 RA Gielens J., Villarroel R., Weitzenger T., Pohl T.M., Terryn N.,
 RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel A.,
 RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
 RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
 RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
 RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
 RA Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C.,
 RA Dalseny M., Voet M., Volckaert G., Meves H.W., Klosternan S.,
 RA Schueller C., Chaiwatiz N.
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 RT Arabidopsis thaliana".
 RL Nature 391:485-488(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EG Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR ENBL: Z97337; CAB10256.1; -
 DR EMBL: AL161539; CAB78519.1; -
 DR InterPro: IPR005172; CXG.
 DR Pfam: PF03638; CXG; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 658 AA; 72129 MW; 5BB99148E5E13A77 CRC64;

Query Match 14.5%; Score 232; DB 10; Length 658;
 Best Local Similarity 39.5%; Pred. No. 8.6e-16;
 Matches 51; Conservative 11; Mismatches 43; Indels 24; Gaps 5;
 YQ 82 PSPALQ-----GPKITLSSGDCFSGGDFC-NSCSC-----NNLRELERFRAKACLD 131
 DB 354 PVEPALQELNLSPPKK---KSYCECFANGVYCIPEPCIDCFNPKPHEDVVLATRKQIES 410

QY 132 RNPEAFQPMKGRGLGA-----AKLRHSGCKNCRSGCLNKYCEYAKIMCSC 180
 DB 411 RNPEAFAPKVRNSDSVQETGDDASKTPASARHKGCKNCKSKLKYCEYGGVGC 470
 QY 181 IKCKIACKN 189
 DB 471 NCRCEGCKN 479

RESULT 13
 Q9M679 PRELIMINARY; PRT; 609 AA.
 AC Q9M679;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE CXG domain containing TSOL-like protein 1.
 GN SOL1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LER;
 RX MEDLINE=20233842; PubMed=10769245;
 RA Hauser B.A., He J., Park S.O., Gasser C.S.;
 RT "TSOL is a novel protein that modulates cytokinesis and cell expansion
 RL in Arabidopsis".
 DR EMBL: AF205142; AAF69125.1; -
 DR InterPro: IPR005172; CXG.
 DR Pfam: PF03638; CXG; 2.
 SQ SEQUENCE 609 AA; 66738 MW; 7E29C055A2423089 CRC64;

Query Match 14.0%; Score 223.5; DB 10; Length 609;
 Best Local Similarity 31.7%; Pred. No. 6.2e-15;
 Matches 52; Conservative 21; Mismatches 36; Indels 55; Gaps 7;
 YQ 98 YDCFSGGDFC-NSCSCNNLRH-----LERFKAICLDNRNPEAFQPMKCGK----- 144
 DB 341 YCECFAGTCTDPCACOGCLNRPVEVETVETKQIESRNPFAFKIVQPTTIDISSHM 397
 YQ 145 RLQ-----AAKLRHSGCKNCRSGCLNKYCEYAKIMCSCICKIACKN----- 189
 DB 398 EVGEDASKTPASARHKGCKNCKSKLKYCEYGGVGCINCRCEGCKNAFGRKDGSL 457
 QY 190 YEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNQ 233
 DB 458 FEQDEENET-----SGTPGTGKTKQ 477

RESULT 14
 Q94AI2 PRELIMINARY; PRT; 526 AA.
 AC Q94AI2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative DNA binding protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Gibson H.A.,
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Yu G., Yu S., Bowser L., Carninci P., Chen H.,
 RA Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,

```
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full length cDNA of gene MW123.15/AT3g22780 (GI:9279696).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046019; AAK76693.1; -
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
SQ SEQUENCE 526 AA; 57886 MW; C6CD9684D417A9D9 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 526;
Best Local Similarity 31.8%; Pred. No. 7.5e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YCDGSSGDFC-NSCSC----NNLRHELRFKAKCLDRNPEAFQPKM----- 141
DB 245 YCEFAAGVYCIIPCSCIDCFNPKIHEETVLATRKQIESRNPFAFAPKVI RNADSIMEAS 304
QY 142 GKRLGAALKRSHKSGCKRSGCLKNYCEYEAKEIMCSICKCIACKN----- 189
DB 305 DDASKTPASARHKGCKNCKKCMKKYCEYQGVGCGSMNCRCEGCTNVFGRKDGSLVI 364
QY 190 ----YESPE--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPPKLRKNR 232
DB 365 MESKLEENQETYEKRIAKIOHNVESKEVQNPSSDQSTPLPPYRHLVHVHQPFLSKNR 423

RESULT 15
Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CXC domain protein TS01 (Putative DNA binding protein).
GN TS01.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LER;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TS01 is a novel protein that modulates cytokinesis and cell expansion
RL in Arabidopsis.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song J.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TS01, an
RL Arabidopsis gene with cysteine-rich repeats.";
RN [3]
RX MEDLINE=1272207; PubMed=12217(2000).
DR EMBL: AF204059; AAF69124.1; -
DR EMBL: AF206324; AAF27433.1; -
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA8550F6E5D4 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 695;
Best Local Similarity 31.8%; Pred. No. 1.1e-14;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YCDGSSGDFC-NSCSC----NNLRHELRFKAKCLDRNPEAFQPKM----- 141
DB 414 YCEFAAGVYCIIPCSCIDCFNPKIHEETVLATRKQIESRNPFAFAPKVI RNADSIMEAS 473
QY 142 GKRLGAALKRSHKSGCKRSGCLKNYCEYEAKEIMCSICKCIACKN----- 189
```

```
Db 474 DDASKTPASARHKGCKNCKKCMKKYCEYQGVGCGSMNCRCEGCTNVFGRKDGSLVI 533
QY 190 ----YESPE--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPPKLRKNR 232
Db 534 MESKLEENQETYEKRIAKIOHNVESKEVQNPSSDQSTPLPPYRHLVHVHQPFLSKNR 592

Search completed: April 21, 2003, 11:49:57
Job time : 29.8114 secs
```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:30:23 ; Search time 33.2222 Seconds
(without alignments)
1199.256 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 NVICQLKGGTQMLCIDNSRT.....GRCLSQLHTFKSKGLKWE 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1647	100.0	299	21 AAY68464	Human testis speci
2	1251	76.0	295	21 AAY68463	Mouse testis speci
3	470	28.5	950	22 ABB62035	Drosophila melanog
4	446	27.1	147	22 ABB96025	Human testicular a
5	446	27.1	147	22 AAM95330	Human reproductive
6	441	26.8	438	20 AAW83392	Caenorhabditis ele
7	378	23.0	280	22 AAG17958	Novel human diagno
8	278	16.9	53	22 AAG76158	Human colon cancer
9	216	13.1	251	22 AAM93348	Human polypeptide,
10	200	12.1	243	22 ABB68888	Drosophila melanog

11	113	6.9	1981	22 ABB61657	Drosophila melanog
12	109.5	6.6	699	20 AAW81840	Human LFA-1 beta c
13	109.5	6.6	769	9 AAF80836	Beta subunit of hu
14	109.5	6.6	769	11 AAR07113	Recombinant beta-s
15	109.5	6.6	769	16 AAR80108	LFA-1 beta subunit
16	109.5	6.6	769	21 AAB03974	LFA-1 CD18 polyep
17	109.5	6.6	793	22 ABG16399	Novel human diagno
18	108.5	6.6	769	13 AAR24256	Beta-subunit CD18
19	108	6.6	676	18 AAW35853	Human CD18 for use
20	108	6.6	3011	13 AAR21519	Compilled HCV seque
21	106.5	6.5	3635	23 ABB81589	Mouse laminin alph
22	106.5	6.5	3635	23 ABB81589	Mouse laminin-15 a
23	105.5	6.4	695	22 ABB30979	Peptide #3630 enco
24	105.5	6.4	695	22 ABB36170	Peptide #3676 enco
25	105.5	6.4	695	22 AAM56953	Human brain expres
26	105.5	6.4	695	22 AAM69338	Human bone marrow
27	105.5	6.4	695	22 AAM17170	Peptide #3604 enco
28	105.5	6.4	695	22 AAM29661	Peptide #3698 enco
29	105.5	6.4	695	22 AAM04863	Peptide #3545 enco
30	105.5	6.4	695	23 ABB38956	Human peptide enco
31	104	6.3	3011	14 AAR31621	Hepatitis C virus
32	103.5	6.3	1725	21 AAB19800	Mouse laminin 2 ma
33	103.5	6.3	1725	21 AAB48451	Mouse laminin 8 po
34	103.5	6.3	1725	23 ABB81593	Mouse laminin 10 s
35	103.5	6.3	1764	10 AAP91672	Primary amino acid
36	103.5	6.3	1776	19 AAW50894	Mouse laminin BI c
37	103.5	6.3	1786	21 AAB19799	Mouse laminin 2 be
38	103.5	6.3	1786	21 AAB48450	Mouse laminin 8 po
39	103.5	6.3	1786	23 ABB81592	Mouse laminin 10 s
40	102	6.2	1788	22 ABB62995	Drosophila melanog
41	101.5	6.2	686	23 AAE18689	HCV-1 NS3/4a mutan
42	101.5	6.2	686	23 AAE18689	Hepatitis C virus
43	101.5	6.2	728	23 AAE18688	NS3/4a mutant conf
44	101.5	6.2	762	22 ABB08032	Novel human diagno
45	101	6.1	473	23 AAM50076	HCV NS3 helicase p

ALIGNMENTS

RESULT 1
AAY68464
ID AAY68464 standard; Protein; 299 AA.
XX
AC AAY68464;
XX
DT 25-APR-2000 (first entry)
XX
DE Human testis specific factor tesmin SEQ ID NO:5.
XX
DE Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
OS Homo sapiens.
XX
PN WO200004147-A1.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-JP03859.
XX
PR 17-JUL-1998; 98JP-0219856.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX
PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX WPI; 2000-147785/13.
DR N-PSDB; AAZ88156.
XX
PT New male germ cell regulatory factor tesmin expressed in spermatocytes

PT useful for investigation of germ cell differentiation and sterility -
 XX Claim 1; Page 50-52; 63pp; Japanese.
 XX The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX Sequence 299 AA;
 SQ

Query Match 100.0%; Score 1647; DB 21; Length 299;
 Best Local Similarity 100.0%; Pred. No. 2.8e-141;
 Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDQNNYLSQSDVPKPMALVGRFLPASTK 60
 Db 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDQNNYLSQSDVPKPMALVGRFLPASTK 60
 Oy 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 120
 Db 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 120
 Oy 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVKNPQHNGKNCRRSGCLKNYCEYEAQI 180
 Db 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVKNPQHNGKNCRRSGCLKNYCEYEAQI 180
 Oy 181 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240
 Db 181 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240
 Oy 241 CISWEVVEATCACLAAOGEAEKHCSCLAEQMILEEFGRLCSQILHTEFKSKGLKME 299
 Db 241 CISWEVVEATCACLAAOGEAEKHCSCLAEQMILEEFGRLCSQILHTEFKSKGLKME 299

RESULT 2
 AAY68463
 ID AAY68463 standard; Protein: 295 AA.
 XX
 AC AAY68463;
 XX
 XX 25-APR-2000 (first entry)
 XX
 XX Mouse testis specific factor tesmin SEQ ID NO:4.
 XX
 XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility.
 XX
 XX Mus musculus.
 XX
 XX WO200004147-A1.
 XX
 XX 27-JAN-2000.
 XX
 XX 16-JUL-1999; 99WO-JP03859.
 XX
 XX 17-JUL-1998; 98JP-0219856.
 XX
 XX (CHUGAI) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX
 XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX
 XX WPI; 2000-147785/13.
 DR N-PSDB; AAZ88155, AAZ88157.
 XX
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 XX Claim 1; Page 47-49; 63pp; Japanese.
 XX

CC The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX Sequence 295 AA;
 SQ

Query Match 76.0%; Score 1251; DB 21; Length 295;
 Best Local Similarity 76.3%; Pred. No. 2.5e-105;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Oy 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDQNNYLSQSDVPKPMALVGRFLPASTK 60
 Db 1 MVICOLKGGTOMLCIDNSRTRELKALHLVPOYQDQNNYLSQSDVPKPMALVGRFLPASTK 60
 Oy 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 120
 Db 61 LNLITQOLEGALPSVNGSAFPGSTLPGPKITLAGYCDGFCASGDFCNCNCCNNL 116
 Oy 121 HHDIERFAIKACLRNPEAFQPKIGKQGLGNVKNPQHNGKNCRRSGCLKNYCEYEAQI 180
 Db 117 RHELEERFAIKACLRNPEAFQPKIGKQGLGNVKNPQHNGKNCRRSGCLKNYCEYEAQI 176
 Oy 181 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 240
 Db 177 MCSSTCKICGCKNYEESPERKTLMSMPNYMOTGGLEGSHYLPPTKFSGLPRFSDRRPSS 236
 Oy 241 CISWEVVEATCACLAAOGEAEKHCSCLAEQMILEEFGRLCSQILHTEFKSKGLKME 299
 Db 237 CISWEVVEATCACLAAOGEAEKHCSCLAEQMILEEFGRLCSQILHTEFKSKGLKME 295

RESULT 3
 ABB62035
 ID ABB62035 standard; Protein: 950 AA.
 XX
 AC ABB62035;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster polypeptide SEQ ID NO 12897.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 KW Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL06138.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of

[illegible]

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483232/52.

Nucleic acids encoding 973 human testicular antigen polypeptides,
useful for preventing, diagnosing and/or treating testicular cancer

Claim 11; SEQ ID NO 1409; 766pp; English.

The present invention provides the protein and coding sequences of 973
human testicular antigens, and fragments of their genomic sequences. The
sequences can be used in the treatment of cardiovascular, urinary system,
reproductive system, immune, respiratory, neurological and
gastrointestinal disorders, infections, and particularly cancer,
especially testicular cancers. The present sequence is a protein of the
invention.

Sequence 147 AA;

Query Match 27.1%; Score 446; DB 22; Length 147;
Best Local Similarity 98.9%; Pred. No. 1.3e-32;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVICQLKGGTQMLCIDNSTRRLKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
|||||
Db 60 MVICQLKGGTQMLCIDNSTRRLKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 119
|||||
QY 61 LNLITQOLEGALPSVYVNGSAFPGSGTLP 88
|||||
Db 120 LNLITQOLEGALPSVYVNGSAFPGSGTLP 147
|||||
RESULT 5
AAM95330
ID AAM95330 standard; Protein; 147 AA.
XX
AC AAM95330;
XX
DT 21-NOV-2001 (first entry)
XX Human reproductive system related antigen SEQ ID NO: 3988.
DE Human reproductive system related antigen; reproductive system disorder;
KW Human; reproductive system related antigen; cancer; gene therapy.
KW
XX Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

ID AAG76158 standard; Protein; 53 AA.
XX
AC AAG76158;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6922.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 11.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR N-PSDB; AAH35563.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 8365; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 53 AA;
Query Match 16.9%; Score 278; DB 22; Length 53;
Best Local Similarity 100.0%; Pred. No. 6.6e-18;
Matches 53; Conservative: 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 VENTACLLAQGEAEKHCSCKLARQMIIEFGRCLSQILHTEFKSKGLKME 299
Db 1 VENTACLLAQGEAEKHCSCKLARQMIIEFGRCLSQILHTEFKSKGLKME 53
RESULT 9
AAH93348
ID AAH93348 standard; Protein; 251 AA.
XX
AC AAH93348;
XX
DT 06-NOV-2001 (first entry)
XX
XX

DE Human polypeptide, SEQ ID NO: 2895.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
DR N-PSDB; AAR94268.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' - and 3' -ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 251 AA;
Query Match 13.1%; Score 216; DB 22; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.9e-11;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVTICQLKGGTQMLCIDSRTRELKALHLVLPQYQDQNNYLQS 41
Db 210 MVTICQLKGGTQMLCIDSRTRELKALHLVLPQYQDQNNYLQS 250
RESULT 10
ABB68888
ID ABB68888 standard; Protein; 243 AA.
XX
AC ABB68888;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33456.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.

AA07113 standard; protein; 769 AA.
AA07113;
24-JAN-1991 (first entry).
Recombinant beta-subunit of LFA-1.
Rhinovirus; LFA-1; Mac-1; p150,95; ICAM-1; autoimmune diseases;
graft rejection.
Key Location/Qualifiers
FH Peptide 197..205
AC /label=tryptic peptide L-56A
XX Peptide 258..269
XX /label=tryptic peptide H58
DT Peptide 311..324
XX /label=tryptic peptide P-61
DE Peptide 340..351
XX /label=tryptic peptide H-52
KW Peptide 367..375
XX /label=tryptic peptide L-65
XX Peptide 472..479
XX /label=tryptic peptide P-20
XX Peptide 489..498
XX /label=tryptic peptide P-18
XX Peptide 528..539
XX /label=tryptic peptide L-56b
PN W09010652-A.
XX
XX
PD 20-SEP-1990.
XX
XX 09-MAR-1990; 90WO-US01255.
XX
PR 09-MAR-1990; 90WO-US01255.
XX
XX (DANA-) DANA FARBER CANCER.
XX
XX Springer T, Kishimoto T, Roberts T;
XX WPI; 1990-304991/40.
DR N-PSDB; AAQ06050.
XX
XX Use of recombinant B-sub-unit of human LFA-1 - for treatment of
PT rhino-viral infections
XX
XX Disclosure; fig 1; 2lpp; English.
XX
XX This beta-subunit of the human lymphocyte function-associated
CC glycoprotein (LFA-1) is used, in a compsn, along with the alpha
CC chain of a member of the CD-18 family, for treating or preventing
CC rhinoviral infections. The beta-subunit interacts with ICAM-1
CC and so inhibits cell-viral attachment. The beta-subunits of other
CC human glycoproteins (esp. Mac-1 or p150,95) involved in cellular
CC adhesion can be used also. Antibodies raised against these can be
CC used as assay reagents to detect LFA-1, etc. in tissues, or for
CC purificn. See also AAQ06051-53.
XX
XX Sequence 769 AA;
Query Match 6.6%; Score 109.5; DB 11; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.37;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVICQLKGTOMLCDINSRTELKALHLVPOYQDONNYLOSVPKPMYALVGRFLPASTK 60
DB 436 IVTVQVLFPQCECRCDQSRSL--CH-----GKFLGGICRCRCDTGYIGKNCRCQTQ 486
QY 61 LNLITQOLEGA-----LPSVYNGSAFPGSGTLPGPKKITLAGYCDCEFA----- 103
DB 487 -GRSSQELGSCRDNNSITCGSLGDCVCGQCLHTSDVPG--KLIIYQYCECDTINCER 543

QY 104 -SGDFCNN-----CNCNNCCNNLHHDIERF-----KAIAKACLRNPEAFQPKIGKGLGN 152
DB 544 YNGQVCGGPGRGLCFCKGK--RCHPGFEGSACQCERTTEGL--NPRRVECS-GRGR--- 595
QY 153 VKPOHNGCNCRRSGCLKNYCECYEAQIM-----CSSIC-----KCIGCKNYEESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCOECPCGCPGCKYKISCAECLKEKGP 638
RESULT 15
AAR80108
ID AAR80108 standard; Protein; 769 AA.
XX
AC AAR80108;
XX
DT 04-MAR-1996 (first entry)
XX
DE LFA-1 beta subunit CD18.
XX
KW CD18; leucocyte function-associated antigen-1; LFA-1; Integrin.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..22
AC /label= sig_peptide
XX 449..496
DT /label= Repeat
XX /note= "cysteine-rich repeat"
DE 497..540
XX /label= Repeat
KW /note= "cysteine-rich repeat"
XX 541..581
XX /label= Repeat
XX /note= "cysteine-rich repeat"
XX 582..617
XX /label= repeat
XX /note= "cysteine-rich repeat"
XX 701..723
XX /label= Trans
XX /note= "transmembrane domain"
XX 724..769
XX /label= Cyto
XX /note= "cytoplasmic domain"
PN W09528170-A1.
XX
XX 26-OCT-1995.
XX
XX 19-APR-1995; 95WO-US04886.
XX
XX 19-APR-1994; 94US-0229513.
XX
XX (UNIV) UNIV KANSAS.
XX
XX Benedict S, Chan MA, Siahaan TJ, Tibbetts SA;
XX WPI; 1995-373631/48.
XX
XX Changing peptide reactivity via conjugation with a second peptide
PT causes change in conformation of first peptide, pref. LFA-1 and
PT ICAM-1 functional domain derived peptide(s)
XX
XX Claim 11; Page 30-32; 64pp; English.
XX
XX Functional domains derived from LFA-1 alpha subunit CD11a (AAR80107),
CC LFA-1 beta subunit CD18 (AAR80107) and ICAM-1 CD43 (AAR80110) are used
CC as the basis of short, synthetic peptides (AAR80109, AAR80111-24) that
CC modulate ICAM/LFA binding interaction. CD18 is obtd. from human
CC tonsil.
XX
XX Sequence 769 AA;
QY

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 21, 2003, 11:48:39 ; Search time 13.5909 Seconds
(without alignments)
647.304 Million cell updates/sec

Title: us-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDSRT.....GRLCSQILHTEFKSLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109.5	6.6	769	2	US-08-789-078-1
2	109.5	6.6	769	2	US-08-752-633-1
3	109.5	6.6	769	2	US-08-476-062A-45
4	109.5	6.6	769	2	US-07-728-215-31
5	109.5	6.6	769	4	US-08-938-085A-31
6	109.5	6.6	769	5	PCT-US95-04886-1
7	109.5	6.6	769	5	PCT-US96-01314-45
8	108	6.6	676	3	US-08-630-172-10
9	108	6.6	676	4	US-09-375-419-10
10	108	6.6	739	4	US-08-444-818-148
11	108	6.6	2995	4	US-08-444-818-138
12	101	6.1	465	2	US-08-833-678A-2
13	101	6.1	465	2	US-08-529-169A-2
14	101	6.1	590	4	US-08-850-328-3
15	101	6.1	631	2	US-08-833-678A-1
16	101	6.1	631	4	US-08-529-169A-1
17	101	6.1	1021	1	US-07-910-760-12
18	101	6.1	1021	1	US-08-440-519-12
19	101	6.1	1021	4	US-08-440-549-12
20	101	6.1	1786	4	US-08-444-818-54
21	101	6.1	2261	4	US-08-444-818-66
22	101	6.1	2436	4	US-08-444-818-75
23	101	6.1	2772	4	US-08-444-818-89
24	101	6.1	2894	2	US-08-466-975A-23
25	101	6.1	2894	2	US-08-391-671A-23
26	101	6.1	2894	3	US-08-467-902A-23
27	101	6.1	2894	4	US-09-275-265-23

28	101	6.1	3011	1	US-08-440-103-36
29	101	6.1	3011	1	US-08-440-542-36
30	101	6.1	3011	1	US-07-910-760-10
31	101	6.1	3011	1	US-08-440-519-10
32	101	6.1	3011	1	US-08-231-368-36
33	101	6.1	3011	1	US-08-440-210-36
34	101	6.1	3011	4	US-09-388-874-2
35	101	6.1	3011	4	US-09-046-604-36
36	101	6.1	3011	4	US-08-440-549-10
37	101	6.1	3011	4	US-08-850-328-1
38	97.5	5.9	969	2	US-08-284-941-2
39	97.5	5.9	969	2	US-08-447-642-2
40	97.5	5.9	969	4	US-09-236-503-2
41	97.5	5.9	969	5	PCT-US93-02147A-2
42	97	5.9	2955	2	US-08-443-260-3
43	97	5.9	2955	3	US-08-442-805A-3
44	97	5.9	2955	3	US-08-443-900A-3
45	97	5.9	2955	4	US-08-444-818-124

ALIGNMENTS

RESULT 1
US-08-789-078-1
; Sequence 1, Application US/08789078
; Patent No. 5843885
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08789,078
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,513
; FILING DATE: 19-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: 816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:

```
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label= cyto
OTHER INFORMATION: /note= "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
TITLE: LFA-1 Amino acid sequence (B2) (from human
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
US-08-789-078-1
```

```
Query Match 6.6%; Score 109.5; DB 2; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVICQLKGGTQMLCIDNSRTRELKALHLPQYQDQNNYLOSVPKPKMTALVGRFLPASTK 60
Db 436 IVTVQVLPQCECRDQSRDSL--CH-----GKFLGCGICRCDTGVIGNCECQTQ 486
QY 61 LNLITQOLEG-----LPSVYNGSAFPGSTLPGPKITLAGVCDFA----- 103
Db 487 -GRSQLEGSKRDNNSIICSLGDCVCGCLCHTSDVPG--KLIYQYCECDTNCER 543
QY 104 -SGDFCN-----CNCNMCNLLHDIERF-----RAIKACLGRNPEAFQPKIGKQGLGN 152
Db 544 YNGQVCGPGRGLFCGKC--RCHPFGSACOCERTTECL--NPRVPCS-GRGR--- 595
QY 153 VKPOHNKNCNRSGCLKNKYCEYEAQIM-----CSSIC-----KCIQCKNYEESP 198
Db 596 -----CRC-----NVCECHSGVQLPLCQECPCPCSPGKYISCACBLKFERGP 638
```

```
RESULT 2
US-08-752-633-1
; Sequence 1, Application US/08752633
; Patent No. 5863889
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Stahaan, Teruna
```

```
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,633
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: (816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label= cyto
```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P31 8717
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 31:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 769 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-07-728-215-31

```


OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label- trans
OTHER INFORMATION: /note- "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label- cyto
OTHER INFORMATION: /note- "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
TITLE: LFA-1 Amino acid sequence (B2) (from human
TITLE: tonsil)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
PCT-US95-04886-1

Query Match 6.6%; Score 109.5; DB 5; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVIQKGGTOMLCIDNSRTRELKALHLVPOYQDNNYLQSDVPKPMALVGRFLPASTK 60
DB 436 IVTVQVLPQCECRCDQSRDSL--CH-----GKFLGEGICRCDTGYIGKNCCECQTQ 486
QY 61 LNLITQOLEGA-----LPSVYNGSAFFSGSTLPGPPKITLAGYCDCEFA----- 103
DB 487 -GRSQELEGSCRDNNIIICGLGDCVGCGLCHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCENN-----CNCNCCNNLHHDIERF-----KAIKACILGRNPEAFQPKIGKQLGN 152
DB 544 YNGQVCGGPGRLCGKGC--RCHPFGESACQCERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VKPQHNKGCNRRSGCLNKYCEYEAQIM-----CSSIC-----KIGCKNYEESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCQECPCGPGKGYISCAECLKFEKGP 638

RESULT 7
PCT-US96-01314-45
Sequence 45, Application PC/TUS9601314
GENERAL INFORMATION:
APPLICANT: M. Amin Arnaout
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
TITLE OF INVENTION: ANTAGONISTS
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01314
FILING DATE: 30-JAN-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/380,167
FILING DATE: 30-JAN-95
ATTORNEY/AGENT INFORMATION:
NAME: John W. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/267001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 769
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US96-01314-45
Query Match 6.6%; Score 109.5; DB 5; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.033;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVIQKGGTOMLCIDNSRTRELKALHLVPOYQDNNYLQSDVPKPMALVGRFLPASTK 60
DB 436 IVTVQVLPQCECRCDQSRDSL--CH-----GKFLGEGICRCDTGYIGKNCCECQTQ 486
QY 61 LNLITQOLEGA-----LPSVYNGSAFFSGSTLPGPPKITLAGYCDCEFA----- 103
DB 487 -GRSQELEGSCRDNNIIICGLGDCVGCGLCHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCENN-----CNCNCCNNLHHDIERF-----KAIKACILGRNPEAFQPKIGKQLGN 152
DB 544 YNGQVCGGPGRLCGKGC--RCHPFGESACQCERTTEGCL--NPRRVECS-GRGR--- 595
QY 153 VKPQHNKGCNRRSGCLNKYCEYEAQIM-----CSSIC-----KIGCKNYEESP 198
DB 596 -----CRC-----NVCECHSGYQLPLCQECPCGPGKGYISCAECLKFEKGP 638
RESULT 8
US-08-630-172-10
Sequence 10, Application US/08630172
Patent No. 6060054
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,172
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-172-10

Query Match 6.6%; Score 108; DB 3; Length 676;

Best Local Similarity 21.5%; Pred. No. 0.039;
Matches 51; Conservative 26; Mismatches 86; Indels 74; Gaps 14;

QY 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLOSDVPKPMALVGRFLPASTK 60
Db 414 IVTVQVLPOCECRCDQSDRSL--CH-----GKFLGCGICRCDTQYIGNKCECQTQ 464
QY 61 LNLITQOLEG-----LPSVWNGSAFPSTLPGPKITILAGYCDGCF----- 103
Db 465 -GRSQLESGCRKDNNSIICSLGDCVCGQCLHTSDVPG--KLIYQYCECDTINCER 521
QY 104 -SGDFCINN-----CNCNCCNHLHDIERF-----KAIAKACLRNPEAFQPKIGKGLGN 152
Db 522 YNGQVCGGPGRLGFCGKC--RCHPGFEGSACOCERTTEGCL--NPRRVECS-GRGR--- 573
QY 153 VKPQHNKGCNRRSGCLKNYCEYEAQIM-----CSSICKICIGCKNYEESP 198
Db 574 -----CRC-----NVCECHSGVQLPLCQPCGCPGCGKYISCAECLKFEKGP 615

RESULT 9

US-09-375-419-10
Sequence 10, Application US/09375419
Patent No. 6264950

GENERAL INFORMATION:

APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172.

FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2879-36

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-375-419-10

Query Match 6.6%; Score 108; DB 4; Length 676;

Best Local Similarity 21.5%; Pred. No. 0.039;
Matches 51; Conservative 26; Mismatches 86; Indels 74; Gaps 14;

QY 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLOSDVPKPMALVGRFLPASTK 60
Db 414 IVTVQVLPOCECRCDQSDRSL--CH-----GKFLGCGICRCDTQYIGNKCECQTQ 464
QY 61 LNLITQOLEG-----LPSVWNGSAFPSTLPGPKITILAGYCDGCF----- 103
Db 465 -GRSQLESGCRKDNNSIICSLGDCVCGQCLHTSDVPG--KLIYQYCECDTINCER 521
QY 104 -SGDFCINN-----CNCNCCNHLHDIERF-----KAIAKACLRNPEAFQPKIGKGLGN 152
Db 522 YNGQVCGGPGRLGFCGKC--RCHPGFEGSACOCERTTEGCL--NPRRVECS-GRGR--- 573
QY 153 VKPQHNKGCNRRSGCLKNYCEYEAQIM-----CSSICKICIGCKNYEESP 198
Db 574 -----CRC-----NVCECHSGVQLPLCQPCGCPGCGKYISCAECLKFEKGP 615

RESULT 10

US-08-444-818-148
Sequence 148, Application US/08444818
Patent No. 6150087

GENERAL INFORMATION:

APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANOV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA

COUNTRY: USA

ZIP: 94508-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/403,590

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Harbin, Alisa A.

REGISTRATION NUMBER: 33,895

REFERENCE/DOCKET NUMBER: 0110.002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (508)359-3876

TELEFAX: (508)359-3885

INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:

LENGTH: 739 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-444-818-148

Query Match 6.6%; Score 108; DB 4; Length 739;

Best Local Similarity 24.9%; Pred. No. 0.043;

Matches 56; Conservative 20; Mismatches 79; Indels 70; Gaps 12;

QY 79 SAFP-SGSTLPGPKITILAGYCDGCFAGDFCNCCNCCNNL-----HHDIERFKA 130
Db 233 SVIPTSGDVVVVATDALTGY-----TGDFDSVIDCNTCTQTVDFSLDFTFTIETLP 287

```

QY      228 G--LPRFSDRRPSSCISWVVEATCACLLQAQEAREHCKSL 270
          |   ||           |   |   ||           |   ||
Db      1580 GENLP-----YLWAYQATVCARAQAAPPSPDQMWKCL 1611

RESULT 12
US-08-833-678A-2
; Sequence 2, Application US/08833678A
; Patent No. 5989905
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,678A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/529,169
; FILING DATE: 15-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-833-678A-2

Query Match          6.1%; Score 101; DB 2; Length 465;
Best Local Similarity 24.8%; Pred. No. 0.11;
Matches 56; Conservative 21; Miscellaneous 77; Models 72.
```

Qy	79	SAFP-SGSTLPGPPKITLAGYCDCAFSGDFCNNCNCCNNL-----HHDIERFKA130
Db	232	SVIPTSGDVVVATDALMTGY-----TGDFSDVICDNTCVTQVDFSLDPTFTIETITLP286
Qy	131	KACLGRNPEAFQPKICKGQLG-----NVKPOHNCNCRSGCLKN--YCCEYDAQIMCSSI185
Db	287	QDAVSRQTR--RGRTGRGPGIYRFVAPGE-----RPSGMFDSVLCCEYDA-----331
Qy	186	CKCIGCKNYEESPERKTLMSMPNYMOTGGL-----EG-----SHYLPPTKE226
Db	332	---GCANYELTP-AETTVRLRAYMTNPGLPVQCQDHLFEWEGVFTGLTHIDAHFLSQTK386
Qy	227	SG--LPRFSDRRPSSCISMEVVYEATCACLLAQGEAEKEHCKCL270
Db	387	SGENLP-----YLVAYQATVCARAOAPPPSDMMKCL419

RESULT 13
US-08-529-169A-2
; Sequence 2, Application US/08529169A
; Patent No. 6194140
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/529,169A
; FILING DATE: 15-SEP-1995
; CLASSIFICATION: 4325
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-529-169A-2

Query Match 6.1%; Score 101; DB 4; Length 465;
Best Local Similarity 24.8%; Pred. No. 0.11;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAPP-SGSTLPGPPIKTLAGYCDCAFSGDFCNCCNCCNNL-----HHDIERFKAI 130
Db 232 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNCVTQVDFSLDPTFTIETILP 286
QY 131 KACLRNPEAFQPKIGKQIGL-----NVKQHNKGCNRRSGCLKN--YCECYEAQIMCSSI 185
Db 287 QDAVSRTOR--RGRTGRGKPGIVRFVAPGE-----RPSGMFDSVLCICYDA----- 331
QY 186 CKICGCKNYEESPERKTLMSMPNYMTGGL-----EG-----SHYLPPTKF 226
Db 332 ----GCAYVELTP-AETTVRLRAYMNTPLPVQCDHLEFWEGVFTGLTHDAHFLSQTQK 386
QY 227 SG--LPRFSDRRPSSCISWEVVEATCACLQAAGEAEKEHCKSL 270
Db 387 SGENLP-----YLVAYQATVCARAQAPPPSNDQMKCL 419

RESULT 14
US-08-850-328-3
; Sequence 3, Application US/08850328
; Patent No. 6379886
; GENERAL INFORMATION:
; APPLICANT: TAKAHAMA, Y.

; APPLICANT: SHIRAIISHI, J.
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS
; TITLE OF INVENTION: C VIRUS INFECTION
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,328
; FILING DATE: 02-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mays, Thomas D
; REGISTRATION NUMBER: 34,524
; REFERENCE/DOCKET NUMBER: 32273-20004.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-850-328-3

Query Match 6.1%; Score 101; DB 4; Length 590;
Best Local Similarity 24.8%; Pred. No. 0.15;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAPP-SGSTLPGPPIKTLAGYCDCAFSGDFCNCCNCCNNL-----HHDIERFKAI 130
Db 374 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNCVTQVDFSLDPTFTIETILP 428
QY 131 KACLRNPEAFQPKIGKQIGL-----NVKQHNKGCNRRSGCLKN--YCECYEAQIMCSSI 185
Db 429 QDAVSRTOR--RGRTGRGKPGIVRFVAPGE-----RPSGMFDSVLCICYDA----- 473
QY 186 CKICGCKNYEESPERKTLMSMPNYMTGGL-----EG-----SHYLPPTKF 226
Db 474 ----GCAYVELTP-AETTVRLRAYMNTPLPVQCDHLEFWEGVFTGLTHDAHFLSQTQK 528
QY 227 SG--LPRFSDRRPSSCISWEVVEATCACLQAAGEAEKEHCKSL 270
Db 529 SGENLP-----YLVAYQATVCARAQAPPPSNDQMKCL 561

RESULT 15
US-08-833-678A-1
; Sequence 1, Application US/08833678A
; Patent No. 5989905
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CHIRON CORPORATION
;; STREET: Intellectual Property - R440, P.O. Box 8097
;; CITY: Emeryville
;; STATE: California
;; COUNTRY: U.S.A.
;; ZIP: 94662-8097
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/833,678A
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/529,169
;; FILING DATE: 15-SEP-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Harbin, Alisa A.
;; REGISTRATION NUMBER: 33,895
;; REFERENCE/DOCKET NUMBER: 0100.005
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 923-3274
;; TELEFAX: (510) 655-3542
;; TELEX: n/a
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 631 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-833-678A-1

Query Match 6.1%; Score 101; DB 2; Length 631;
Best Local Similarity 24.8%; Pred. No. 0.17;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAFP-SGSTLPGPPTITLAGYCDGFCNCCNCCNL-----HHDIERFKAI 130
| | | | : | | : | | : | | | |
Db 398 SVIPTSGDVVVVATDALMTGY-----TGDFSDVIDCNTCTQTVDVDFSLDPTFTIETILP 452
| | | | : | | : | | : | | | |
QY 131 KACLGHNPEAFQPKIGKGQLG---NVKQHNKGCNCRRSGLKN--YCFCYEAQIMCSSI 185
: : : : | : | : | | : | | | |
Db 453 QNAVSTQR--RGRTRGRGPGYRFVAPGE-----RPSGMFDSSVLCECYDA----- 497
| | | | : | | : | | : | | | |
QY 186 KCKGCKNYEESPERKTLMSMPNMQTGGI-----EG-----SHYLPPTKF 226
| | | | : | | : | | | |
Db 498 ----GCAYTELTP-AETTVRLRAYMNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSQTKQ 552
| | | | : | | : | | | |
QY 227 SG--LPRFSDRRPSSCISWEVVEATCACLQAQGEAEKEHCKSL 270
| | | | : | | : | | | |
Db 553 SGENLP-----YLVAYQATVCARAQAPPPSWDMWKCL 585

Search completed: April 21, 2003, 11:51:11
Job time : 15.5909 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:49:09 ; Search time 14.5976 Seconds
(without alignments)
1548.781 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRLSQLHTFEKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437	26.5	438	10	US-09-220-091-13
2	112.5	6.8	1192	9	US-10-189-971-18
3	112.5	6.8	1207	9	US-10-189-971-20
4	112.5	6.8	1477	9	US-10-189-971-8
5	112.5	6.8	1535	9	US-10-189-971-14
6	112.5	6.8	1593	9	US-10-189-971-4
7	109.5	6.6	769	9	US-10-072-841-31
8	106.5	6.5	3635	9	US-10-037-182-4
9	106.5	6.5	3635	10	US-09-845-583-2
10	105.5	6.4	695	10	US-09-864-761-42967
11	105	6.4	1057	9	US-10-189-971-6
12	105	6.4	1251	9	US-10-189-971-16
13	105	6.4	1342	9	US-10-189-971-24
14	105	6.4	1512	9	US-10-189-971-10
15	105	6.4	1570	9	US-10-189-971-12
16	105	6.4	1628	9	US-10-189-971-2
17	103.5	6.3	1725	9	US-10-037-182-12
18	103.5	6.3	1786	9	US-10-037-182-10
19	103.5	6.3	1786	10	US-09-938-275-7

20	102.5	6.2	4123	9	US-10-213-509-5
21	101.5	6.2	686	9	US-09-881-239-3
22	101.5	6.2	686	10	US-09-881-654-2
23	101.5	6.2	728	9	US-09-881-239-1
24	101	6.1	1099	10	US-09-881-654-4
25	101	6.1	2894	10	US-10-044-995-23
26	101	6.1	2894	10	US-09-941-611-23
27	101	6.1	3011	10	US-09-916-359-2
28	100	6.1	2743	9	US-10-037-182-36
29	100	6.1	3695	9	US-10-037-182-2
30	99.5	6.0	1551	10	US-09-970-318-4
31	98.5	6.0	2749	9	US-10-123-155-385
32	97.5	5.9	632	10	US-09-929-955-29
33	97.5	5.9	686	10	US-09-929-955-17
34	97.5	5.9	686	10	US-09-929-955-31
35	97.5	5.9	686	10	US-09-929-955-32
36	97.5	5.9	686	10	US-09-929-955-43
37	97.5	5.9	686	10	US-09-929-955-44
38	97.5	5.9	686	10	US-09-929-955-45
39	97.5	5.9	686	10	US-09-929-955-46
40	97.5	5.9	686	10	US-09-929-955-47
41	97.5	5.9	686	10	US-09-929-955-48
42	97.5	5.9	686	10	US-09-929-955-49
43	97.5	5.9	732	9	US-09-977-418-20
44	97.5	5.9	1036	9	US-09-373-967-4
45	97.5	5.9	1036	9	US-10-028-072-142

ALIGNMENTS

RESULT 1
US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Query Match	26.5%	Score 437;	DB 10;	Length 438;
Best Local Similarity	41.3%	Pred No. 2 3e-29;		
Matches	92;	Conservative 35;	Mismatches 62;	Indels 34;
Gaps	8;			
QY	98	YDCFPASGDFCNCCNCCNLLHDIERFKAICACLRNPEAFQPKIG--KGQLGNVXP	155	
Db	191	YDCFPANGFCDNCCKDCHNNIEYDSQRKATQSLRNPNAPKPKIGIARGITDIER	250	
QY	156	QHNKGCNRRSCLKNYCEYEAQIMCSICKICIGCKN-----YES-PRKTL	203	
Db	251	LHQKCHCKSKCLKNYCEYEAQIMCSICKICIGCKN-----YES-PRKTL	310	
QY	204	MSMPNYMQT-----GGL-----EGSHYLPPTKFSGLPRFSDRR--PSSCSISWEVEAT	250	
Db	311	MSLTNASSSTATPDGSGSVVYDEHGGDY-EDMLLSHKPKVEMDPFRFPYMTDEVVEAA	369	
QY	251	CACLLAQGEA-----EKEHCSCKLAEQMILEEFGRCLSQIL	287	

D**b** 370 TMCVQAEEALNYEKVQTEDEKLINMEKLVLREFFGRCLEQMI 412

RESULT 2

```

US-10-189-971-18
; Sequence 18, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: NO. US20030028907A1el
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 1192

```

LENGTH: 1192

```

; ELEMENT: 122
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-18

```

Query Match	Score	DB	Length
6.88	112.5	9	1192

Best Local Similarity 20.08; Predict. No. 0.24; Matches 59; Conservative 14; Mismatches 75; Indels 147; Gaps 14;

Qy	78	GSAPFSGSTLPGP	-----PKTTLAGYCDCFASGDFCNNCN	-----	112
Db	326	GKEFASGERFSP	TAAACHLCLCWEGSVCEPKACAPALCFPPARGDCDPCDCEYLGES	385	
Qy	113	-----CNNCNNLHHDIERFAIKACLG	-----RNPEAFQPKIGKQ	149	
Db	386	YLSNQEPDPREPCNL	-----TCLGFFVTCGRRCPEPCGCSHPPLPSGH	430	
Qy	150	L	-----GNVKP	-----QHNKGCNRRSGCLKNYCEYEA	178
Db	431	CCPTCGCRYHGVTTASGETLPDPDPTCSLCTCGREHQDGEFFEGPAGSCWECRCQAG	490		
Qy	179	QIMC	-----SSIC-KCIGCKNY-BESPERKTLMSMPNYMTQTGGLEGSH	219	
Db	491	QVSCVRLOCPPLCKLQVTERGSCCPRCGLAHGEHP	-----EGSR	533	
Qy	220	YLPPTKFSGLPREFSHDRRSPSCISWEVVENTAC	--CL--LAQGEAEAEKEHGSKC	269	
Db	534	WVPP	-----DSACSSVCYCHEGV-VTCARIOCISSCAPROGPHDPCOC	576	

RESULT 3

```

US/10-189-971-20
; Sequence 20, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20

```

SOFTWARE: F0
SEQ ID NO 20

; LENGTH: 1207
 ; TYPE: PRT
 ; ORGANISM: ho
 US-10-189-971-20

Query Match	6.8%;	Score 112.5;	DB 9;	Length 1207;
Best Local Similarity	20.0%;	Pred. No. 0.25;		
Matches	59: Conservative	14: Mismatches	75: Indels	147: Gaps

Qy	78	GSAPPSGSTLPGP-----PKITLAGYCDCFASGDFCNNCN-----	112
Db	341	GKEFASGERFPSTAACHLCLCWGWSVSCFKACAPALCFPPARGDCDCDCEYLGES	400
Qy	113	-----CNCNCNLHDIERFKAIAKCLG-----RNPEAFQPKIGKQ	149
Db	401	YLSNQEFDPREPCNL-----TCLGFFVTCGRRCPCPCGCSHPLIPSGH	445
Qy	150	L-----GNVKP-----QHNKGNCRRSGCLKNKYCECYBA	178
Db	446	CCPTCOGRYHGVTASGETLPDLPDPTCSLCTCGREHODGEFEGPAGSCWECRCQAG	505
Qy	179	QIMC-----SSIC-KCIGCKNV-BESPERKLTMSMNPYMTGGLEGSH	219
Db	506	QVSCVRQLCPPLCKLQVTERGSCPCRCRCLANGHEHP-----EGSR	548
Qy	220	YLPTKFSGLPRFSDHRRPSSCISWEVEATCA---CL--LAQGEAEKEHCSCK	269
Db	549	WVPB-----DSACSSVCHEGV-VTCARIOCISSCAOPRGPHDCCPOC	591

RESULT 4

```

US-10-189-971-8
; Publication 8, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander J
; TITLE OF INVENTION: No. US20030028
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/30
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/31
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Vers
; SEQ ID NO 8
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-8

```

Query Match 6-88: Score 112.5: DB 9: Length 1477:

	Query Match	Score 112.76, E-Val 9e-38, Length 14777,
	Best Local Similarity	20.08; Pred. No. 0.31;
	Matches 59; Conservative	14; Mismatches 75; Indels 147; Gaps 14
Qy	78 GSAPFSSTLPGP-----PKITLAGYCDCFASGDGCNNCN-----	112
Db	611 GKFAAGERFSPTAACHLCILCWEGSVCEPKACAPALCPARGDCDDCGCEYLGES	670
Qy	113 -----CNCNCCNLHDIERKAIKACLG-----RNPEAFQPIKGKG	149
Db	671 YLSNQEFDPREPCLC-----TCLGGFVTCRRRPECPEGGSPLIPSGH	715
Qy	150 L-----GNVKP-----OHNGKCNRSSGLKNKYCEYE	178
Db	716 CCPTQCGRYHGVTASGETLPDLPDTCSLTCTQGREGDEEFGAGSCWRCQAG	775
Qy	179 QIMC-----SSIC-KIGCKNY-EESPERKTLMSPMNYMTGGLESH	219

UNIT 1

RESULT 14

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:20:36 ; Search time 210.79 seconds
(without alignments)
499.275 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDSRT.....GRLSOLHTFKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main :*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1647	100.0	299	21	US-09-743-237-5
2	1251	76.0	295	21	Sequence 5, Appli
3	1251	76.0	295	21	Sequence 4, Appli
4	533.5	32.4	223	1	Sequence 24, Appl
5	470	28.5	403	26	PCT-US01-08656-10693
6	470	28.5	890	26	Sequence 10693, A
7	470	28.5	950	20	Sequence 1700, Ap
					Sequence 12948, A
					Sequence 12897, A

8	470	28.5	950	26	US-60-191-637-12935	Sequence 12935, A
9	446	27.1	147	1	PCT-US01-01329-1409	Sequence 1409, Ap
10	446	27.1	147	1	PCT-US01-01339-3988	Sequence 3988, Ap
11	446	27.1	147	21	US-09-764-891-3988	Sequence 3988, Ap
12	446	27.1	147	24	US-10-080-090-1409	Sequence 1409, Ap
13	437	26.5	438	14	US-09-087-136-13	Sequence 13, Appl
14	437	26.5	438	14	US-09-087-136-13	Sequence 13, Appl
15	437	26.5	438	16	US-09-220-091-13	Sequence 13, Appl
16	378	23.0	280	1	PCT-US01-08631-48317	Sequence 48317, A
17	376	22.8	524	21	US-09-708-427-33367	Sequence 33367, A
18	376	22.8	550	21	US-09-708-427-33366	Sequence 33366, A
19	376	22.8	571	21	US-09-708-427-33365	Sequence 33365, A
20	369	22.4	615	21	US-09-733-089-22002	Sequence 22002, A
21	369	22.4	615	21	US-09-733-089-22183	Sequence 22183, A
22	369	22.4	615	21	US-09-733-089-22184	Sequence 22184, A
23	369	22.4	615	22	US-09-816-660-22002	Sequence 22002, A
24	369	22.4	615	22	US-09-816-660-22183	Sequence 22183, A
25	369	22.4	615	22	US-09-816-660-22184	Sequence 22184, A
26	366.5	22.3	241	26	US-60-324-109-27382	Sequence 27382, A
27	357	21.7	603	19	US-09-573-655A-2154	Sequence 2154, Ap
28	357	21.7	603	19	US-09-573-655A-2192	Sequence 2192, Ap
29	357	21.7	603	21	US-09-708-427-27294	Sequence 27294, A
30	353	21.4	534	21	US-09-708-427-27295	Sequence 27295, A
31	351	21.3	518	21	US-09-708-427-27296	Sequence 27296, A
32	316.5	19.2	207	21	US-09-733-089-22013	Sequence 22013, A
33	316.5	19.2	207	22	US-09-816-660-22013	Sequence 22013, A
34	314.5	19.1	168	21	US-09-733-089-21980	Sequence 21980, A
35	314.5	19.1	168	22	US-09-816-660-21980	Sequence 21980, A
36	278	16.9	53	1	PCT-US00-26524B-6922	Sequence 6922, Ap
37	263	16.0	349	18	US-09-428-944-1881	Sequence 1881, Ap
38	257.5	15.6	386	26	US-60-324-109-16915	Sequence 16915, A
39	257.5	15.6	556	26	US-60-324-109-32886	Sequence 32886, A
40	245.5	14.9	497	21	US-09-708-427-25631	Sequence 25631, A
41	245.5	14.9	542	21	US-09-708-427-25630	Sequence 25630, A
42	245.5	14.9	658	21	US-09-708-427-25629	Sequence 25629, A
43	245	14.9	770	26	US-60-324-109-17053	Sequence 17053, A
44	245	14.9	770	26	US-60-324-109-17063	Sequence 17063, A
45	242.5	14.7	402	26	US-60-324-109-31233	Sequence 31233, A

ALIGNMENTS

RESULT 1
US-09-743-237-5
; Sequence 5, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YOUNJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-743-237-5

Query Match 100.0%; Score 1647; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 6.5e-148;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVICQLKGGTQMLCIDSRTRELKALHLVPOYQDNQNNYLOSQDPKPM TALVGRFLPASTK 60
|||||
DB 1 MVICQLKGGTQMLCIDSRTRELKALHLVPOYQDNQNNYLOSQDPKPM TALVGRFLPASTK 60

[illegible]

RESULT 11
US-09-764-891-3988
; Sequence 3988, Application US/09764891
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3988
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3988

Query Match 27.1%; Score 446; DB 21; Length 147;
Best Local Similarity 98.9%; Pred. No. 6.9e-34;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MYICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYQLQSDVPKMTALVGRFLPASTK 60
Db 60 MYICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYQLQSDVPKMTALVGRFLPASTK 119
QY 61 LNLITQLEGALPSVYVNGSAPPSGSLP 88
Db 120 LNLITQLEGALPSVYVNGSAPPSGSTXP 147

RESULT 12
US-10-080-090-1409
; Sequence 1409, Application US/10080090
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL20C1
; CURRENT APPLICATION NUMBER: US/10/080,090
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 3506
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1409
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-080-090-1409

Query Match 27.1%; Score 446; DB 24; Length 147;
Best Local Similarity 98.9%; Pred. No. 6.9e-34;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MYICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYQLQSDVPKMTALVGRFLPASTK 60
Db 60 MYICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYQLQSDVPKMTALVGRFLPASTK 119
QY 61 LNLITQLEGALPSVYVNGSAPPSGSLP 88
Db 120 LNLITQLEGALPSVYVNGSAPPSGSTXP 147

RESULT 13
US-09-087-136-13
; Sequence 13, Application US/09087136
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

Query Match 26.5%; Score 437; DB 14; Length 438;
Best Local Similarity 41.3%; Pred. No. 2.2e-32;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;
QY 98 YDCDFASGDFCNCCNCCNLLHDIERFKAICACLAGRNPEAFOPKIG--KGOLGNVKP 155
Db 191 YDCDFANGFEFCDCKCKDCHNNIYDSQSKALROSLEARNPNAFKPKIGIARGGITDIER 250
QY 156 QHKGNCNRRSGCLKNYCEYEAQIMCSICKICGCKN-----YEES----PERKTL 203
Db 251 LHQKGCHCKKSGCLKNYCEYEAQVPTDRCKCKGQNTETRYMTRYKNSGGAVSNTNAL 310
QY 204 MSMPNMQT-----GGL-----EGSHYLPPTKFSGLPRFSDRR--PSSCISWEVVEAT 250
Db 311 MSLTNASSTATPDSPGVSVTDEHGGDY-EDMLLSHKPKVENDPRFRFPYMYMTDEVVEAA 369
QY 251 CACLLAQGEA-----EKECHSKCLAEOMILEEFGRCLSQIL 287
Db 370 TMCWVAQAEALNYEKVQTEDEKLNMEKLVLEFRGRCLEQMI 412

RESULT 14
US-09-087-136-13
; Sequence 13, Application US/09087136A
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136A
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

Query Match 26.5%; Score 437; DB 14; Length 438;
Best Local Similarity 41.3%; Pred. No. 2.2e-32;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;
QY 98 YDCDFASGDFCNCCNCCNLLHDIERFKAICACLAGRNPEAFOPKIG--KGOLGNVKP 155
Db 191 YDCDFANGFEFCDCKCKDCHNNIYDSQSKALROSLEARNPNAFKPKIGIARGGITDIER 250
QY 156 QHKGNCNRRSGCLKNYCEYEAQIMCSICKICGCKN-----YEES----PERKTL 203
Db 251 LHQKGCHCKKSGCLKNYCEYEAQVPTDRCKCKGQNTETRYMTRYKNSGGAVSNTNAL 310

[illegible]

RESULT 15

```

US-09-220-091-13
; Sequence 13, Application US/09220091
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/720,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

```

[illegible]

Search completed: July 11, 2002, 08:20:36
Job time: 420 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:14:15 ; Search time 30.19 Seconds
(without alignments)
1028.575 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVTCQLKGQTQMLCIDNSRT.....GRCLSQLHTFRKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 323636 seqs, 103855142 residues
Total number of hits satisfying chosen parameters: 323636

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	366.5	22.3	282	6	US-10-155-881-9626
2	357	21.7	603	5	US-09-573-655B-2154
3	357	21.7	603	5	US-09-573-655B-2192
4	348	21.1	510	6	US-10-155-881-9875
5	278	16.9	53	6	US-10-105-698-6932
6	248.5	15.1	896	6	US-10-155-881-20031
7	201.5	12.2	359	6	US-10-155-881-9183
8	183	11.1	143	6	US-10-155-881-9877
9	167	10.1	120	6	US-10-138-145-1127
10	112	6.8	1774	1	PCT-US02-01339-6
11	109.5	6.6	699	1	PCT-US02-09671-718
12	109.5	6.6	761	1	PCT-US02-09671-711
13	109.5	6.6	769	1	PCT-US02-09671-708
14	109.5	6.6	769	1	PCT-US02-09671-710
15	109.5	6.6	769	1	PCT-US02-09671-712
16	109.5	6.6	769	1	PCT-US02-09671-713
17	109.5	6.6	769	1	PCT-US02-09671-714
18	109.5	6.6	769	1	PCT-US02-09671-715
19	109.5	6.6	769	1	PCT-US02-09671-716
20	109.5	6.6	769	1	PCT-US02-09671-719
21	109.5	6.6	769	1	PCT-US02-09671-720
22	109.5	6.6	769	6	US-10-072-844-31
23	109.5	6.6	769	6	US-10-072-838-31
24	109.5	6.6	769	6	US-10-099-007A-12
25	109.5	6.6	817	6	US-10-144-793-141
26	108	6.6	676	1	PCT-US02-09671-717

ALIGNMENTS

RESULT 1
US-10-155-881-9626
; Sequence 9626, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyva, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9626
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9626

Query Match 22.3%; Score 366.5; DB 6; Length 282;
Best Local Similarity 47.7%; Pred. No. 1.3e-23;
Matches 63; Conservative 22; Mismatches 26; Indels 21; Gaps 3;
QY 98 YDCFCASGDFCNCCNCCNLLHDDIERFKAICACGLGNPEAFQPKIGKQL---GNVK 154
Db 82 YCECFASGYDCGCTNCFNPNNEARVEAELDRNPDAFRPKIGSSPLMHRNEV 141
QY 155 P-----OHKGCNCRSSGCKNRYCEYEAQIMCSICKICGCKNYEESPERKTLMSMP 207
Db 142 PSDLPVAGKHNGCHCKKSGCLKKYCECFQANTLCSENCKDCKNFESEERNL---- 197
QY 208 NYMQTGGLEGSH 219
Db 198 -----FQGDH 202

RESULT 2
US-09-573-655B-2154
; Sequence 2154, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18

; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6932

Query Match 16.9%; Score 278; DB 6; Length 53;
Best Local Similarity 100.0%; Pred. No. 6.3e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 247 VEATCACLAAQGEAEKEKCSKLAEOIMLEEFGRCLSOILHTEPKSKGLKME 299
Db 1 VEATCACLAAQGEAEKEKCSKLAEOIMLEEFGRCLSOILHTEPKSKGLKME 53

RESULT 6
US-10-155-881-20031
; Sequence 20031, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 20031
; LENGTH: 896
; TYPE: PRT
; ORGANISM: Glycine max
US-10-155-881-20031

Query Match 15.1%; Score 248.5; DB 6; Length 896;
Best Local Similarity 40.7%; Pred. No. 5.3e-13;
Matches 50; Conservative 13; Mismatches 31; Indels 29; Gaps 5;
QY 98 YDCDFASGDFCNCN-CNCNCCNLHHDIERFKAIKACL-GRNPEAFQPKIGKQGLGNVXP 155
Db 487 YDCDFAAAGTYCTDPCACQGLNRPEY-VETVVETKQIESRNPFIAPAKI-----VQP 538
QY 156 -----QHNGKNCRRSGCLKNYCEYEAQIMCSSICKIGCKNYEE 196
Db 539 TTDISSHMDENLTPSSARHRCNCNRCMCLKKYCEYQANVGSSGCRCEGCKNVHG 598
QY 197 SPE 199
Db 599 KKE 601

RESULT 7
US-10-155-881-9183
; Sequence 9183, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9183
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9183

Query Match 12.2%; Score 201.5; DB 6; Length 359;
Best Local Similarity 23.1%; Pred. No. 1.8e-09;
Matches 63; Conservative 35; Mismatches 102; Indels 73; Gaps 9;
QY 98 YDCDFASGDFCNCN-CNCNCCN-----NLHHD-----IERFKAIKACLGRNPEAFQPKIG 146
Db 80 YPCFSGGGYCDKCGCQCFNKEAFETVHTRKVLLSRQKMSMKINRREANTEPME 139
QY 147 KQQLGNVQPHNKGNCRRSGCLKNYCEYEAQIMCSSICKIGCKNYEESPERKTLMSM 206
Db 140 DAHSSSSSTPPKRGCGNCKSKLKKYCDYQDGTGCSLFCRCDDCONPFGKNEGIMADDS 199
QY 207 PNYMOTGG---LEGS-----HYLPP-----TK 225
Db 200 KRYLTGADLDHSEGEHDFVVERSPLOSPIKSESFHOTPPHLRASSRDAHVFPHAISQ 259
QY 226 FSGLPFRSH--DRPSSCIS-----WEVVE-----ATCACLAAQGEAEKE 264
Db 260 WQALPRSWHCNKNRSDRAMODSANYKNSHDYQVQPKHEDSYSISKVCQILNGMAELS 319
QY 265 HCSKCLAEOMILEEFGRCLSQILHTEFKSKGLK 297
Db 320 QVEKSVAPDVFQPGNREIFVSLSGDVRAMWLK 352

RESULT 8
US-10-155-881-9877
; Sequence 9877, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9877
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9877

Query Match 11.1%; Score 183; DB 6; Length 143;
Best Local Similarity 40.9%; Pred. No. 2.4e-08;
Matches 36; Conservative 16; Mismatches 24; Indels 12; Gaps 3;
QY 98 YDCDFASGDFCNCNCCNCCNLHHDIERFKAIKACLGRNPEAFQPKIGK-GQLGNV---- 153
Db 55 YCECFQELQYCDGCGNCGNIVGNENARNEAIEAIRQNPISAFQPKIGNDSNTLNVKRD 114
QY 154 -----KPQHNKGNCRRSGCLKNYCE 174
Db 115 KFWSSPSRSKPHKCHCKKS-CLKKYCD 141

RESULT 9
US-10-138-145-1127
; Sequence 1127, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen

APPLICANT: Shenk, Michael Andrew
APPLICANT: Gibson, John Bryan
APPLICANT: Norriss, Michael Geoffrey
APPLICANT: Saulsbury, Keith Martin
APPLICANT: Hall, Claire
TITLE OF INVENTION: Compositions isolated from forage
FILE REFERENCE: 11000.1058U
CURRENT FILING DATE: 2002-05-01
NUMBER OF SEQ ID NOS: 1552
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1127
LENGTH: 120
TYPE: PRT
ORGANISM: Lolium perenne
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(120)
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-138-145-1127

Query Match 10.1%; Score 167; DB 6; Length 120;
Best Local Similarity 37.4%; Pred. No. 4.4e-07;
Matches 40; Conservative 15; Mismatches 36; Indels 16; Gaps 8;
QY 98 YDCDFASGDFCNN-CNCCNCLNHHDIERFAKACACLRNPEAFOPKI-----GKGLGN 152
Db 16 YCEE-AAGVYCEPCGCGCLNRPHEIV-LSTRKQIEFRNPLA-APKVRLSDAAQETQ 72
QY 153 VKP-----QHNKGCNRRSGCKLNKCYEYEAQIMCSCICKICGCKN 193
Db 73 EDNTPASARHKRCNCKKSSCLK-YCEYGGVGLTNCRC-ECKN 117

RESULT 10
PCT-US02-01339-6
Sequence 6, Application PC/TUS0201339
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: LEE, Ernestine A.
APPLICANT: WALIA, Narinder K.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: AZIMZAI, Yalda
APPLICANT: TANG, Y. Tom
APPLICANT: YUE, Henry
APPLICANT: THANGAVELU, Kavitha
APPLICANT: XU, Yuming
APPLICANT: ARVIZU, Chandra
APPLICANT: WARREN, Bridget A.
APPLICANT: YAO, Monique G.
APPLICANT: AU-YOUNG, Janice
APPLICANT: HAFALIA, April J.A.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: KALLICK, Deborah A.
APPLICANT: GANDHI, Ameena R.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: KHAN, Farrah A.
APPLICANT: LU, Yan
APPLICANT: SWARNAKAR, Anita
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: NGUYEN, Dannel B.
APPLICANT: GRAUL, Richard
APPLICANT: LU, Dyung Aina M.
TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
FILE REFERENCE: PI-0346 PCT
CURRENT FILING DATE: 2002-01-16
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PERL Program

SEQ ID NO 6
LENGTH: 1774
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No: 7722591CDI
PCT-US02-01339-6

Query Match 6.8%; Score 112; DB 1; Length 1774;
Best Local Similarity 20.2%; Pred. No. 0.46;
Matches 51; Conservative 33; Mismatches 73; Indels 96; Gaps 15;
QY 86 TLPGPKITLAGYCDGFCNN---CN-CNNCC---NNLH-----H 122
Db 800 TLPDP-----CRLLSPEACNQSACTWCHGACLSGDQAHRLGGGSPSPMPSPE 852
QY 123 DIERFAKACLRNPEAFOPKIGKQLGNVKNPKHNGCNCRSGCL-----KNYCE 174
Db 853 ECRRLRTCSECLARHPRTLQP--GDGEAST--PRCKWCTNCEGACIGRNGSCTSEDCR 908
QY 175 CYEAQIMCSCIC-----KICCKNYEESPERKTLMSM-PNY----- 209
Db 909 INOREVFWAGNCSEAAAGAACDCEQCTREGKCMWTRQFKRTGTRRLSVQPTDYDTFCFSH 968
QY 210 ----MOTGGLEGSHYLP-PTKFSGLPRFSDRRPSSCISWEVVEATCACLGAEEAEKE 264
Db 969 SLNVSPMPVESFPPLPCPTPCHLLP-----NCTS-----CLDSKAGDGGWQ 1010
QY 265 HC-----SKCLA 271
Db 1011 HCVWSSSLQOCLS 1023

RESULT 11
PCT-US02-09671-718
Sequence 718, Application PC/TUS0209671
GENERAL INFORMATION:
APPLICANT: ZYCOS Inc.
TITLE OF INVENTION: TRANSLATIONAL PROFILING
FILE REFERENCE: 08191-026W01
CURRENT FILING DATE: 2002-03-28
CURRENT APPLICATION NUMBER: PCT/US02/09671
PRIOR FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: 60/279,495
PRIOR FILING DATE: 2001-03-28
PRIOR APPLICATION NUMBER: 60/292,544
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: 60/310,801
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/326,370
PRIOR FILING DATE: 2001-10-01
PRIOR APPLICATION NUMBER: 60/336,780
PRIOR FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 60/358,985
PRIOR FILING DATE: 2002-02-20
NUMBER OF SEQ ID NOS: 2041
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 718
LENGTH: 699
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-09671-718

Query Match 6.6%; Score 109.5; DB 1; Length 699;
Best Local Similarity 22.3%; Pred. No. 0.26;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVTCOLKGGTQMLCINDSRTRKALHLVPOYQDNNYLOSVPKPM TALVGRFLPASTK 60
Db 436 IVTVQVLPOCECRCDQRDRSL--CH-----GKGFLECGICRCDTYIGNKCECQTQ 486

Job time: 40 sec

Best Local Similarity 22.3%; Pred. No. 0.29;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

1	QY	MYTCQLKGTOMLCTDNSTREKALHLVPOQDQNNYLSQSDPKPMWALVGRFLPASTK	60
436	Db	YVLPVPOQECRCRQSDRSJL--CH-----GKGFLECGICRCDTGYIGKNCECQTQ	486
61	QY	LNLTITQOLEGA-----LPSVYNGSAFPSSGSLTPGPPKTLTACYGDCDFA----	103
487	Db	-GRSQLEGGSGCKDNNSIICSLGDCVCGQCLCHTSDVPG--KLIIYGYCECDTINCR	543
104	QY	-SGDFCNN-----CNCNCCNLLHHDIERF-----KAIKACLGRNPEAFQPKRIGKQLGN	152
544	Db	YNGQVCGGPGRLGFCGKC--RCHPFGESACOCERTTEGCL--NPRRVBCS-GRGR----	595
153	QY	VKPQHNGKNCRRSGCLKNWCBCYEAQIM-----GSSIC-----KCIKGKNYEESP	198
596	Db	-----CRC-----NVCECHSGYUPLQCEQCPGSPCGKYISCAECLKFKSGP	638

RESULT 15

```

PCT-US02-09671-712
; Sequence 712, Application PC/TUS0209671
; GENERAL INFORMATION:
; APPLICANT: zygos Inc.
; TITLE OF INVENTION: TRANSLATIONAL PROFILING
; FILE REFERENCE: 08191-026W01
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/335,780
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 712
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671-712

```

[illegible]

Search completed: July 11, 2002, 08:14:16

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 11:45:13 ; Search time 15.6044 Seconds
(without alignments)
1842.060 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQLHTEFSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437	26.5	429	2 T23152	hypothetical prote
2	376	22.8	571	2 B84585	hypothetical prote
3	357	21.7	603	2 T08955	hypothetical prote
4	245.5	14.9	658	2 F71410	hypothetical prote
5	117	7.1	497	2 T27827	hypothetical prote
6	112	6.8	1737	2 T00209	MEGF8 protein - hu
7	109.5	6.6	769	1 L3HULM	leukocyte adhesion
8	108.5	6.5	3635	2 T10053	laminin alpha 5 ch
9	105.5	6.4	2180	2 T29784	hypothetical prote
10	104	6.3	775	2 T21436	hypothetical prote
11	103.5	6.3	1786	1 MNMSB1	laminin beta-1 cha
12	102	6.2	1790	1 MNFEB1	laminin beta-1 cha
13	101	6.1	1700	2 S08167	Balbani ring 3 pr
14	101	6.1	1819	2 A71928	cag island protein
15	101	6.1	3011	1 GNWVC3	genome polyprotein
16	100.5	6.1	633	2 T24898	hypothetical prote
17	100	6.1	395	2 S4454	transcription fact
18	100	6.1	520	2 G84510	probable receptor-
19	99.5	6.0	677	2 C42125	trophozoite cystei
20	99	6.0	373	2 T47488	hypothetical prote
21	99	6.0	265	2 T16408	hypothetical prote
22	99	6.0	2195	2 T34264	hypothetical prote
23	98.5	6.0	574	2 B88465	protein B0244.8 [i
24	98.5	6.0	962	2 JC5571	subtilisin-like pr
25	98.5	6.0	3106	1 S53868	laminin alpha-2 ch
26	98.5	6.0	5376	2 T42215	zonadhesin - mouse
27	97.5	5.9	432	2 T37509	hypothetical prote
28	97.5	5.9	969	1 A39490	subtilisin-like pr
29	97.5	5.9	975	2 JC5570	subtilisin-like pr

30	97.5	5.9	1766	2 A42125	trophozoite cystei
31	97	5.9	382	2 T29339	hypothetical prote
32	97	5.9	798	2 A40526	integrin beta-7 ch
33	97	5.9	3010	1 A45573	genome polyprotein
34	96.5	5.9	290	2 G72858	AcOrf-70 protein -
35	96.5	5.9	1895	2 T15881	hypothetical prote
36	96.5	5.9	3759	2 A35085	trithorax protein
37	96	5.8	798	2 A28193	integrin beta-1 ch
38	96	5.8	3010	1 GNWVC3	genome polyprotein
39	95.5	5.8	1801	1 MNRTS	laminin beta-2 cha
40	95	5.8	1927	2 G64585	cag pathogenicity
41	94	5.7	3011	1 GNWVC3	genome polyprotein
42	93.5	5.7	772	2 S32659	integrin beta 2 ch
43	93.5	5.7	1599	2 T16210	hypothetical prote
44	93	5.6	1751	1 MNHUMH	laminin alpha-2 ch
45	93	5.6	1810	1 A32230	tenascin precursor

ALIGNMENTS

RESULT 1

T23152

hypothetical protein JC8.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23152

R:Lightning, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19699

A:Accession: T23152

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-429 <WIL>

A:Cross-references: EMBL:Z82274; PIDN:CA805228.1; GSPDB:GN00022; CESP:JC8.6b

A:Experimental source: clone JC8

C:Genetics:

A:Gene: CESP:JC8.6b

A:Map position: 4

A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Query Match 26.5%; Score 437; DB 2; Length 429;

Best Local Similarity 41.3%; Pred. No. 6.9e-28;

Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;

Qy 98 YDCFCASGDFCNCCNCCNHLHDIERFKAKACLRNPEAFQPKIG--KGQLGNVKP 155

Db 182 YDCFCANGFCRDCNCKDCHNIEYDSQSKAIRSLERNPNAPKPKIGIARGGITDIER 241

Qy 156 QHNKGCNRRSGCLKNYCEYEAQIMCSSICKICCKN-----YEES---PERKTL 203

Db 242 LHQKCHCKKSGCLKNYCEYEAQVPTDRCKCKCQNTETRYMTRYKNSGGAVENTNAL 301

Qy 204 MSMPNYMT-----GGL-----EGSHYLPPTKFGSLPRFSDHRR--PSSCISWVEAT 250

Db 302 MSLTNASTATPDSGFSVVTDEHGDY-EDMLLSHKPKVEMDPRFRFYMTDEVVEAA 360

Qy 251 CACLLAQGEA-----EKEHSCKCLAEQMILEEFGRLCSQL 287

Db 361 TMCNVAQAEEALNKEVKQVTEDEKLNMEKLVLEFRGRCLEQMI 403

RESULT 2

B84585

hypothetical protein At2g20110 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: B84585

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A02870; MUID:85051302; PMID:6209134

A;Accession: A02871

A:Molecule type: mRNA

A;Residues: 1292-1530,'MEMP','1535-1691','C','1693-1748','N','1750-1786'<BAR>

R;tross-references: EMBL:X05122; NID:g52861; PIDN:CRAA28839.1; PID:g809042

R;Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbauer, I.; Hartl, L.

Eur. J. Biochem. 177, 35-45, 1988

A:title: Structural study of long arm fragments of laminin. Evidence for repetitive C

A;Reference number: S01790; MUID:89030693; PMID:3181157

A;Accession: S02036

A:Molecule type: protein

A;Residues: 1561-1587 <DEU>

R;Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.

EMBO J. 4, 309-316, 1985

A:title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.

A;Reference number: S13543; MUID:85257455; PMID:3848400

A;Accession: S13543

A:Molecule type: protein

A;Residues: 1700-1748,'N','1750-1759'<PAU>

C;Genetics:

A:Gene: Lamb-1

A:Map position: 12

C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin

C;Function:

A:description: Interact with cells and with other basement membrane proteins to promote superfamily: laminin beta-1 chain; laminin-type EGF-like homology

C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular

F;-21/Domain: signal sequence #status predicted <SIG>

F;22-1786/Product: laminin beta-1 chain #status predicted <MAT>

F;22-270/Domain: VI <DOM6>

F;271-540/Domain: V <DOM5>

F;271-332/Domain: laminin-type EGF-like homology <LE01>

F;335-395/Domain: laminin-type EGF-like homology <LE02>

F;398-455/Domain: laminin-type EGF-like homology <LE03>

F;458-507/Domain: laminin-type EGF-like homology <LE04>

F;510-540/Domain: laminin-type EGF-like homology <LE05>

F;541-772/Domain: IV <DOM4>

F;773-1182/Domain: III <DOM3>

F;773-818/Domain: laminin-type EGF-like homology <LE06>

F;821-864/Domain: laminin-type EGF-like homology <LE07>

F;867-914/Domain: laminin-type EGF-like homology <LE08>

F;917-973/Domain: laminin-type EGF-like homology <LE09>

F;976-1025/Domain: laminin-type EGF-like homology <LE10>

F;1028-1081/Domain: laminin-type EGF-like homology <LE11>

F;1084-1129/Domain: laminin-type EGF-like homology <LE12>

F;1132-1176/Domain: laminin-type EGF-like homology <LE13>

F;1183-1397/Domain: II <DOM2>

F;1183-1397/Region: heptad repeats

F;1398-1430/Domain: alpha <ALP>

F;1431-1786/Region: heptad repeats

F;1431-1786/Domain: I <DOM1>

F;22/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;320-356/Disulfide bonds: #status predicted

F;120,356,519,677,1041,1195,1279,1336,1343,1487,1533,1542,1643/Binding site: carbonyd

F;1179,1182,1785/Disulfide bonds: Interchain #status predicted

Query Match 6.3%; Score 103.5; DB 1; Length 1786;
Best Local Similarity 21.4%; Pred.No. 2.9;
Matches 54; Conservative 29; Mismatches 74; Indels 95; Gaps 15;

QY 90 PPKITLAGYCDCFASGDFFCNCC-----NCNNCCNNHLLHDIERFKAIKACLR 136
I : I : I : I : I :
Db 937 PVTLQLACVDCPYIGSRDDCASGFGFNPDGSGQCPC--QCHNIIDT-----T 985
I : I : I : I : I :

QY 137 NPFAFPQTKG-----GQLGNVPQINKGCNRRRSCLKNY-----C 173
I : I : I : I : I :
Db 986 DPBACDKDTGRCLKLYHTEGDHQLCQCYGYDGLRDQRCKVCNYLGTVREHCNGSDC 1045
I : I : I : I : I :

QY 174 ECYEAOIMCSSICKICCKNYEESPERRKTLMSPNTWQ---TG-----GLESHYLPTT 224
I : I : I : I : I :
Db 1046 HCDKATGTSCSLPNVIIG-----QNCRD----CAPNTWLASGTGGCPGCCNAHSFPGSC 1096
I : I : I : I : I :
QY 225 -KFSG----LPRESHRRRPSSC--ISW-----EVVEATCACLLAA 257
I : I : I : I : I :

Db 1097 NEFTGQCQMFPG-GRTSCSEQLFWGDPDVECRACDPRGIETPQCDSQGCVCVE 1155
QY 258 GEEAEKEHCKSC 269
Db 1156 GVEGPR-CDKC 1165

RESULT 12

MMFVB1 laminin beta-1 chain precursor - fruit fly (Drosophila melanogaster)

N;Alternate names: laminin chain B1
C;Species: Drosophila melanogaster

C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C;Accession: A28783; S14462; B28783

R;Montell, D.J.; Goodman, C.S.

Cell 53, 463-473, 1998

A;Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals do

A;Reference number: A28783; MUID:88210471; PMID:3365769

A;Accession: A28783

A;Molecule type: mRNA

A;Residues: 1-1790 <MON1>

A;Cross-references: EMBL:M19525

R;Montell, D.J.; Goodman, C.S.

submitted to the EMBL Data Library, June 1988

A;Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reve

A;Reference number: S14462

A;Accession: S14462

A;Molecule type: mRNA

A;Residues: 1-667, 'L', 669-725, 'VT', 728-947, 950-1790 <MON2>

A;Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802

C;Genetics:

A;Gene: lamB1

A;Cross-references: FlyBase:FBgn0002527

A;Map position: 2L 28D

C;Function: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin

C;Superfamily: interact with cells and with other basement membrane proteins to promote

C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular

F;1-26/Domain: signal sequence #status predicted <SIG>

F;27-1790/Product: laminin beta-1 chain #status predicted <MAT>

F;27-288/Domain: VI <DOM6>

F;289-561/Domain: V <DOM5>

F;290-354/Domain: laminin-type EGF-like homology <LE01>

F;357-417/Domain: laminin-type EGF-like homology <LE02>

F;420-477/Domain: laminin-type EGF-like homology <LE03>

F;480-528/Domain: laminin-type EGF-like homology <LE04>

F;531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>

F;562-789/Domain: IV <DOM4>

F;643-645/Region: cell attachment (R-G-D) motif

F;790-1189/Domain: III <DOM3>

F;791-836/Domain: laminin-type EGF-like homology <LE06>

F;839-882/Domain: laminin-type EGF-like homology <LE07>

F;885-932/Domain: laminin-type EGF-like homology <LE08>

F;935-990/Domain: laminin-type EGF-like homology <LE09>

F;968-972/Region: cell adhesion #status predicted

F;993-1042/Domain: laminin-type EGF-like homology <LE10>

F;1045-1093/Domain: laminin-type EGF-like homology <LE11>

F;1096-1141/Domain: laminin-type EGF-like homology <LE12>

F;1144-1188/Domain: laminin-type EGF-like homology <LE13>

F;1190-1407/Domain: II <DOM2>

F;1408-1434/Domain: alpha <ALP>

F;1435-1790/Domain: I <DOM1>

F;51-56/Disulfide bonds: #status predicted

F;140,203,234,489,593,1053,1248,1303,1332,1343,1475,1495,1517,1583,1646,1705/Binding sit

F;1191,1194,1788/Disulfide bonds: interchain #status predicted

Query Match 6.2%; Score 102; DB 1; Length 1790;

Best Local Similarity 21.4%; Pred. No. 3.8;

Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

QY 97 GYCD-----CF--ASGDFCNCC-----NCCNC-----CNLLHDIERFKA 129

Db 1005 GNCDRQTGACLKLYQTGTGHCLECKDGFEGDALQNCQCECDFLGTNTNTIAHCDRTG 1064
QY 130 IKACLRGNPEAFOPKIGKGGOLGNVKKPOHNK-----GNCRRSGCLKNVCEYEAQIM 181
Db 1065 QCPCL-----PNVQGVRCDOCAENHWKIASGEGESCNCDDPICALHEQCNSTYGCQ 1116
QY 182 CS---SICKCIGCK-NYEESEPKTLMSMPNYMTGGLEGSHYLPPTKFS--GLPRFSHD 235
Db 1117 CKPGFGGRACNQCAHYWGNPK-----CQCECDQFGAADFQCD 1157
QY 236 RRRSSCISWEVETATCCLLAQGEAEKEHCKSLAQOMILEEFGRLS 284
Db 1158 RETGNCVCHGIGGYKNECARGYIGQPHRCSPC-----GECFN 1196

RESULT 13

S08167

Balbani ring 3 protein - midge (Chironomus tentans)

C;Species: Chironomus tentans

C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 21-Jul-2000

C;Accession: S08167

R;Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.

J. Mol. Biol. 211, 331-349, 1990

A;Title: The balbani ring 3 gene in Chironomus tentans has a diverged repetitive str

A;Reference number: S08167; MUID:90172404; PMID:1689777

A;Accession: S08167

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-1700 <PAU>

A;Cross-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058

C;Genetics:

A;Gene: BR3

A;Map position: 4

C;Superfamily: unassigned Balbani ring proteins

Query Match 6.1%; Score 101; DB 2; Length 1700;

Best Local Similarity 22.6%; Pred. No. 4.4;

Matches 51; Conservative 20; Mismatches 87; Indels 68; Gaps 15;

QY 99 CDCFASGD-----FC-NCCNC-----NCCNLLHDIERFKAICLGRNEAF 141

Db 198 CSCECKGDGKCGSKIWCKNCRICPTAPAGGCSAPLKWDDD--KCSACAPARMEKK 255

QY 142 QPKIGKQGLGNVKNKNGCNCRRSGCLKN-----YCEYE-----AQIMCSICK 187

Db 256 EKVESKIKWN---PNTCEGCAQLNCPDNKKANKETCCQCEKVKCKNGGVFCCKDCS 312

QY 188 CIGCKNYEESPERKTLMSMPNYMTGGLEGSHYLP---PTKFSGLPR----- 231

Db 313 CV-CPGGDKD---KTCTAPQVY---DGVACSCSPVNNQKPADGCPRPQKWDKECRCEC 365

QY 232 -FSHRRPSSCISWEVVEATCAC-----LLAQGEAEKEHCKS-KCL 270

Db 366 PVKHDCKNGKV--WDETICICIPRDAPVCTACKERGCECKCI 409

RESULT 14

A71928

cag island protein - Helicobacter pylori (strain J99)

C;Species: Helicobacter pylori

A;Variety: strain J99

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C;Accession: A71928

R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.

Nature 397, 176-180, 1999

A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A;Reference number: A71800; MUID:99120557; PMID:9923682

A;Accession: A71928

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1819 <ARN>

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1251	76.0	475	11	Q8VIE0	mus musculus
2	1244	75.5	475	11	Q8VIE1	mus musculus
3	470	28.5	950	5	Q3V608	drosohilla
4	437	26.5	429	5	O6Z295	caenorhabdi
5	437	26.5	435	5	Q35Q07	caenorhabdi
6	376	22.8	571	10	Q9SL70	arabidopsi
7	369	22.4	277	11	Q9DS71	mus musculu
8	357	21.7	603	10	Q9SZD1	arabidopsi
9	262.5	15.9	356	10	Q9LW71	arabidopsi
10	261	15.8	601	10	Q9CAV1	arabidopsi
11	248.5	15.1	896	10	Q9Z522	glycine max
12	247.5	13.0	609	10	Q9M079	arabidopsi
13	245.5	14.9	658	10	O23333	arabidopsi
14	229.5	13.9	526	10	Q94A12	arabidopsi
15	229.5	13.9	695	10	Q9LE32	arabidopsi
16	229.5	13.9	695	10	Q9LUI3	arabidopsi

Query Match	22.4%	Score 369;	DB 11;	Length 277;				
Best Local Similarity	73.2%	Pred. No. 3.1e-30;						
Matches	71;	Conservative	8;	Mismatches 18; Indels 0; Gaps				
QY 1	MVTCOLKGGTQMLCIDNSRTRKALHLVPOYODNNYLOSQVDPKMTALVGRFLPASTK 60							
Db 181	MVTCOLKGGAGMCLIDNCGAREKALHLPPQYDQSFQSELPKMTTLVGRLLPVPK 240							
QY 61	LNLTQOEGALPSVNGSAFFPGSTLPGPKITLAG 97							
Db 241	LNLTQVDNGALPSVNGAFAFGPALQGPKITLSC 277							
RESULT 8								
Q9SZD1	PRELIMINARY; PRT; 603 AA.							
AC Q9SZD1								
DT 01-MAY-2000	(TEMBLrel. 13, Created)							
DT 01-JUN-2000	(TEMBLrel. 13, Last sequence update)							
DT 01-JUN-2002	(TEMBLrel. 21, Last annotation update)							
DE	Putative transcription factor.							
GN F19B15.30	OR AT4G29000.							
OS Arabidopsis thaliana	(Mouse-ear cross).							
OC Eukaryota;	Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;								
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi..								
OC NCBI_TaxID=3702;								
RN [1]								
RP SEQUENCE FROM N.A.								
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,								
RA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;								
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.								
RN [2]								
RP SEQUENCE FROM N.A.								
RA EU Arabidopsis sequencing project.								
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.								
RN [3]								
RP SEQUENCE FROM N.A.								
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,								
RA Mewes H.W., Lemcke K., Mayer K.F.X.;								
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.								
RN [4]								
RP SEQUENCE FROM N.A.								
RA EU Arabidopsis sequencing project;								
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.								
DR EMBL; AL078470; CAB43914.1; -								
DR EMBL; AL161574; CAB79658.1; -								
DR InterPro; IPR005172; CXC.								
DR Pfam; PF03638; CXC; 2.								
SQ SEQUENCE 603 AA; 64635 MW; 7FD1951AFB818DE5 CRC64;								
Query Match								
Best Local Similarity 21.7%; Score 357; DB 10; Length 603;								
Matches 96; Conservative 43; Mismatches 83; Indels 118; Gaps								
QY 39	LQSDVP---KP-MTALVGRF-----LPASTKUNLTQOEGALPSVNGSAFFPGSGSTLP 88							
Db 58	ISSVSTPIRPGMTIAIGQVTVRPTLPMTATNPNPSSQ-----IVNA---PIRHP 109							
QY 89	GPPIK-----TLAGYCDCFASGDFCNCCNCCNNLHH 122							
Db 110	ESPAKGRPNVEGRDGTQKKQCKNCKHSRCLKLYCECFASGTCDGCNCVNCFNVDN 169							
QY 123	DIERFKAIKACLRGNPEAFQPKTKGQGLGN-----VKPQMGKNCRRSGCLKN 171							
Db 170	EPARREAVEATLRNPFAPRPKIASPPGGRDKREDIGEVVLLGKNGKCHCKKSGCLKK 229							
QY 172	YCCEYEAQINCSICKICGCKNVEESPRTKLSMPN-----YMQ-----TGGLEGS 219							
Db 230	YCFQANILCSNCLDCKNPFEGSEERQALFHGEHSNHMYLQQAANAITGAVGSSG 289							
QY 220	YLP-----PTFSGLPFRFSDRR---PSS----- 240							

DE Putative DNA binding protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Gibson H.A.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yu G., Yu S., Bowser L., Carninci P., Chen H.,
RA Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MW123.15/AT3g22780 (GI:9279696).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY046019; AAK76693.1; -;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 536 AA; 57886 MW; C6CD9684D417A9D9 CRC64;

Query Match 13.9%; Score 229.5; DB 10; Length 526;
Best Local Similarity 24.2%; Pred. No. 2.7e-15;
Matches 55; Conservative 31; Mismatches 72; Indels 69; Gaps 5;

Qy 98 YDCDFASGDFC-NNCNCNNCCNHLHDIERFKATKACLRNPEAFQPKI----- 145
Db 245 YCECFAAGVYCEPCSCIDCFNKPIHEETVLATRKQIESRNPLAPAPKVI RNADSIMEAS 304
Qy 146 GKGQLGNVYKPOHNGKNCRRSGCLKNYCECYEAQIMCSICKICGCKNY----- 194
Db 305 DDASKTPASARHKRCNCCKSNCKMKKYCECYGGVGCSSMNCRCGCTNVFGKDKGSLVI 364
Qy 195 -----EESPERKTLMSMPNYMOTGGL-----GSHYL 221
Db 365 MESKLEENQETYEKRIAKIOHNVEVSKEVEQNPPSSDQSTPLPPYRHLVWHQPFSLKNRL 424
Qy 222 PPTKRF-----SGLPRFSDRRPSSCISWEVVEATCACL 255
Db 425 PPTQFFLGTGSSFRKPNSDLAQSQNEKKPLETVTDKTEIMPEILL 471

RESULT 15
Q9LE32
ID Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CXC domain protein TS01 (Putative DNA binding protein).
GN TS01.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. LER;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TS01 is a novel protein that modulates cytokinesis and cell expansion
in Arabidopsis.";
RL Development 127:2219-2226(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song J.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TS01, an
Arabidopsis gene with cysteine-rich repeats.";

RL Development 127:2207-2217(2000).
DR EMBL; AF204059; AAF69124.1; -;
DR EMBL; AF206324; AAF27433.1; -;
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA8550F6E5D4 CRC64;

Query Match 13.9%; Score 229.5; DB 10; Length 695;
Best Local Similarity 24.2%; Pred. No. 3.8e-15;
Matches 55; Conservative 31; Mismatches 72; Indels 69; Gaps 5;

Qy 98 YDCDFASGDFC-NNCNCNNCCNHLHDIERFKATKACLRNPEAFQPKI----- 145
Db 414 YCECFAAGVYCEPCSCIDCFNKPIHEETVLATRKQIESRNPLAPAPKVI RNADSIMEAS 473
Qy 146 GKGQLGNVYKPOHNGKNCRRSGCLKNYCECYEAQIMCSICKICGCKNY----- 194
Db 474 DDASKTPASARHKRCNCCKSNCKMKKYCECYGGVGCSSMNCRCGCTNVFGKDKGSLVI 533
Qy 195 -----EESPERKTLMSMPNYMOTGGL-----GSHYL 221
Db 534 MESKLEENQETYEKRIAKIOHNVEVSKEVEQNPPSSDQSTPLPPYRHLVWHQPFSLKNRL 593
Qy 222 PPTKRF-----SGLPRFSDRRPSSCISWEVVEATCACL 255
Db 594 PPTQFFLGTGSSFRKPNSDLAQSQNEKKPLETVTDKTEIMPEILL 640

Search completed: April 21, 2003, 11:49:59
Job time : 30.1886 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:19:52 ; Search time 95.879 Seconds
(without alignments)
6228.992 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 4008
Sequence: 1 tctcctgtggtggcccg.....aaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09743237/runat_21042003_111943_381/app_query.fasta_1.4494
-DB=A_Geneseq_101002 -QFAST=FASTAN -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -LOCAL -OUTFMT=ptc -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 -ACGN_1_1_259 -runat_21042003_111943_381 -NCPUP=6 -ICPU=3
-NO_XLPYX -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1599	39.9	295	21	AAV68463	Mouse testis speci
2	1251	31.2	299	21	AAV68464	Human testis speci
3	503	12.5	147	22	AB96025	Human testicular a
4	503	12.5	147	22	AA95330	Human reproductive
5	440.5	11.0	950	22	AB62035	Drosophila melanog
6	405.5	10.1	438	20	AA83392	Caenorhabditis ele
7	362	9.0	280	22	ABG17958	Novel human diago
8	346	8.6	251	22	AA931348	Human polypeptide
9	248	6.2	53	22	AA931348	Human colon cancer
10	186	4.6	243	22	AB68888	Drosophila melanog
11	140.5	3.5	4561	22	ABG30203	Novel human diago
12	140.5	3.5	9222	22	ABG21064	Novel human diago
13	136.5	3.4	966	23	AA80189	Human TSP1 domain
c 14	131.5	3.2	1061	20	AA87504	Human N-methyl-D-a
c 15	127	3.1	961	23	AA80188	Human TSP1 domain
16	126.5	3.2	1454	22	AA40222	Human polypeptide
17	125	3.1	961	23	AA80188	Human TSP1 domain
c 18	124.5	3.1	1081	21	AA826240	Human N-methyl-D-a
c 19	124.5	3.1	1081	23	AB856523	Human NM2A recepto
c 20	124	3.1	3680	22	AB70878	Drosophila melanog
21	122.5	3.1	1017	22	AB85813	TutD protein #4
22	122.5	3.1	1615	22	AB85826	Protein #3 encoded
c 23	120.5	3.0	966	23	AA80189	Human TSP1 domain
c 24	120.5	3.0	1212	20	AA87503	Human N-methyl-D-a
c 25	119.5	3.0	386	22	AA69494	Human purified sec
c 26	119.5	3.0	1185	20	AA93497	Human atrophin I p
c 27	119.5	3.0	2594	16	AAW14748	IgG-Fc binding pro
c 28	119.5	3.0	3541	23	AA85130	Human melanoma spe
c 29	119.5	3.0	5405	16	AAW14749	IgG-Fc binding pro
c 30	119.5	3.0	7337	22	ABG22216	Novel human diago
c 31	118.5	2.9	1413	23	AAE21729	Human PKIN-24 prot
32	118.5	3.0	1422	22	ABG02345	Novel human diago
33	118.5	3.0	3118	22	AB850362	Human SRCAP. Homo
34	117.5	2.9	732	18	AAW26642	Human RECK cancer-
35	117.5	2.9	2972	22	AB850363	Human SRCAP. Homo
36	117	2.9	566	22	AA42020	Human polypeptide
c 37	117	2.9	1028	21	AAV68734	A receptor-like pr
c 38	115.5	2.9	1033	22	ABG23220	Novel human diago
c 39	114	2.8	1082	23	AB90349	Human polypeptide
c 40	114	2.8	2157	21	AA93910	A human hyaluronan
c 41	113.5	2.8	1232	21	AB826239	Human N-methyl-D-a
c 42	113.5	2.8	1232	23	AA47961	Human NM2A recepto
43	113.5	2.8	2971	21	AA841231	Human ORFX ORF995
44	113	2.8	349	21	AAV44855	Mouse taste cell s
45	113	2.8	746	22	AB859201	Drosophila melanog

ALIGNMENTS

RESULT 1
AAV68463

ID AAV68463 standard; Protein; 295 AA.

XX AAV68463;

AC AAV68463;

XX 25-APR-2000 (first entry)

DE Mouse testis specific factor tesmin SEQ ID NO:4.
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.

OS Mus musculus.

XX WO200004147-A1.

XX 27-JAN-2000.

PD

XX 16-JUL-1999; 99WO-JP03859.
PF
XX 17-JUL-1998; 98JP-0219856.
PR
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX
XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX WPI: 2000-147785/13.
DR N-PSDB; AA288155; AA288157.
XX
XX New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility -
XX
XX Claim 1; Page 47-49; 63pp; Japanese.
XX
XX The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin. Tesmin
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
XX Sequence 295 AA;
SQ
Alignment Scores:
Pred. No.: 2,79e-142 Length: 295
Score: 1599.00 Matches: 295
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39,90% Indels: 0
DB: 21 Gaps: 0
US-09-743-237-1 (1-2241) x AAY68463 (1-295)
QY 651 ATGGTATTTGTAGCTGAAGAGGGCCGAGTCTGTGCATAGACACTGTGGCGG 710
DB 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 711 AGGAGCTCAAGCGCTCCATCTGCTCCCTCAGTACGATCACCAGACGACTTCCCTCAG 770
DB 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerPheProGln 40
QY 771 TCAGAGCTCCCTAAGCCCAATGACAACTTTAGTGGGAGAGACTTCTGCCAGTACCAGCGAAG 830
DB 41 SerGluLeuProLysProMetThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 831 TTAATCTCATCACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCATGGGGCTGCC 890
DB 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 891 TTTCCCTCTGGACTGCTCTGCAAGGGCCACCACCAAAATAACTCTGTCTGGGTACTGTGAC 950
DB 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGCGGGGACTTCTGCAACAGCTGCGAGCTGCAACACCTGCGCCCATGAGCTC 1010
DB 101 CysPheSerSerGlyAspPheCysAsnSerCysSerCysAsnLeuArgHisGluLeu 120
QY 1011 GAGCGTTCACAGCCATAAAGCGGTGTCTTGATAGAAATCCTGAAGCTTTCCCAACCAAAA 1070
DB 121 GluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheGlnProLys 140
QY 1071 ATGGGAAAGCGGCTCTGGAGCTGTAACTTCGACACAGCAAGGCTGCAACTGTAAG 1130
DB 141 MetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGlyCysAsnCysLys 160
QY 1131 CGCTCAGGCTGCCGTAAGAACTACTGTGAGTGTATGAGGCCAAATCATGTGTCTTCC 1190
DB 161 ArgSerGlyCysLeuLysAsnTyrCysGlyCysTyrGluAlaLysIleMetCysSerSer 180
QY 1191 ATTTGCAATGCATGTGCTTCCAAAACTATGACAAAGTCCAGAACCAAAATGCTGATG 1250
|||||

Db 181 IleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArgLysMetLeuMet 200
QY 1251 AGCACCCACATACATGAGCCTGGGACTTTGAGAGCAGCAGCATTTTGTCCCGAGCC 1310
|||||
Db 201 SerThrProHisTyrMetGluProGlyAspPheGluSerSerHisTyrLeuSerProAla 220
QY 1311 AGTTCTCAGGACCTCCAAAACCTAGAAAATAGGAGGCGCTTCTCCTGTATCTCCTGG 1370
|||||
Db 221 LysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSerCysIleSerTrp 240
QY 1371 GAAGTAGTGAGGCGCACATGTGCTGCTGCGCCAGGCTGAGGAAGCAGAGCAGGAG 1430
|||||
Db 241 GluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluAlaGluGlnGlu 260
QY 1431 CACTGTTCCTCCCAAGCTTGGCTGAGCAGATGATCCTCGAGGAGTTTGGAGGTCCTGTCG 1490
|||||
Db 261 HisCysSerProSerLeuAlaGluGlnMetIleLeuGluGluPheGlyArgCysLeuSer 280
QY 1491 CAGATTCTCCATCAGTTCAGTTCAGTTCAGGGGCTGAAAATTGAG 1535
|||||
Db 281 GlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295
RESULT 2
AAY68464
ID AAY68464 standard; Protein; 299 AA.
XX
AC AAY68464;
DT 25-APR-2000 (first entry)
XX
DE Human testis specific factor tesmin SEQ ID NO:5.
XX
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
OS Homo sapiens.
XX
PN WO200004147-A1.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-JP03859.
XX
PR 17-JUL-1998; 98JP-0219856.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX
XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
XX WPI: 2000-147785/13.
DR N-PSDB; AA288156.
XX
PT New male germ cell regulatory factor tesmin expressed in spermatocytes
XX useful for investigation of germ cell differentiation and sterility -
XX
PS Claim 1; Page 50-52; 63pp; Japanese.
XX
CC The present sequence represents a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin. Tesmin
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.
XX
SQ Sequence 299 AA;
Alignment Scores:
Pred. No.: 2,47e-109 Length: 299
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 31.21% Indels: 4
DB: 21 Gaps: 1

US-09-743-237-1 (1-2241) x AAY68464 (1-299)	
QY 651	ATGGTATTTCTCAGCTGAAGAGGCGGCCAGATGCTCTGCATACACAACTCTGGCGG 710
Db 1	MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
QY 711	AGGAGCTCAAGCGCTCCATCTGCTTCCTCAGTACGATGACACAGACGAGTTTCCCTCAG 770
Db 21	ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGln 40
QY 771	TCAGAGCTCCCTAAGCAATGACAACCTTTAGTGGGAGACTTCTGCCAGTACCAGCGAAG 830
Db 41	SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
QY 831	TTAAATCTCATCACAGCTTGATATGGAGCTCTCCCATCAGCTCTCAATGGGCTGCC 890
Db 61	LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
QY 891	TTTCCCTCTGGAGCTGCTCTGCAAGGGCCACCCAAATAACTCTGTGGTACTGTGAC 950
Db 81	PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
QY 951	TGCTTCTCCAGCGGGACTTCTCCACAGCTGCAGC-----TGCACAACCTG 998
Db 101	CysPheAlaSerGlyAspPheCysAsnAsnCysAsnCysAsnAsnLeu 120
QY 999	CGCATGAGCTCGAGCGCTTCAAGCCATAAGCGCTGCTTGTATAGAACTCTGAAGCT 1058
Db 121	HisHisAspIleGluArgPheLysAlaIleLysAlaCysLeuGlyArgAsnProGluAla 140
QY 1059	TTCCAAACCAAAATGGGAAAGCGCTCTGGAGCTGCTAAACTTCGACACAGCAAGGG 1118
Db 141	PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
QY 1119	TGCAACTGTAAAGCGCTCAGCTGCCGTGAGAACTACTGTGAGTGTATAGGCCAAATC 1178
Db 161	CysAsnCysArgArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGlnIle 180
QY 1179	ATGTGTTCTTCCATTGCAAACTGCTTGCATAAACTATCAAGAACTCCAGCAACGA 1238
Db 181	MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200
QY 1239	AAATGCTGATGACACACCCACTACATGGAGCGCTGGGACCTTTGAGAGCAGCCATTAT 1298
Db 201	LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
QY 1299	TTGTCGCCAGCCAACTTCTCAGCACTCCAAACTGAGAAAATAGGAGCGCTTCTCC 1358
Db 221	LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgProSerSer 240
QY 1359	TGTATCTCTGGGAAGTGTGGAGGCCACATGTCCTGCTGCTGCCCGCCAGGTGAGGAA 1418
Db 241	CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 260
QY 1419	GCAGAGCAGGAGCAGCTTTCCCAAGCTTGGCTGAGCAGATGATCTGGAGGAGTTTGA 1478
Db 261	AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluPheGly 280
QY 1479	AGTGGCTGTGAGATTTCTCCACATCGAGTTCAGTCCAAAGGGGTGAAATTTGAG 1535
Db 281	ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299
RESULT 3	
ABB96025	
ID	ABB96025 standard; Protein: 147 AA.
AC	
XX	ABB96025;
XX	
DT	21-JUN-2002 (first entry)
XX	
DE	Human testicular antigen SEQ ID NO: 1409.
XX	

Human: testicular antigen; testes: cancer; metastasis; immune disorder;
reproductive system disorder; urinary system disorder; gene therapy;
cardiovascular disorder; respiratory disorder; neurological disorder;
gastrointestinal disease; infection; cytostatic.

Homo sapiens.

WO200155317-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US01329.

31-JAN-2000; 2000US-0179065.

04-FEB-2000; 2000US-0180628.

24-FEB-2000; 2000US-0184664.

02-MAR-2000; 2000US-0186350.

16-MAR-2000; 2000US-0189874.

17-MAR-2000; 2000US-0190076.

18-APR-2000; 2000US-0198123.

19-MAY-2000; 2000US-0205515.

07-JUN-2000; 2000US-0209467.

28-JUN-2000; 2000US-0214886.

30-JUN-2000; 2000US-0215135.

07-JUL-2000; 2000US-0216647.

07-JUL-2000; 2000US-0216880.

11-JUL-2000; 2000US-0217487.

11-JUL-2000; 2000US-0217496.

14-JUL-2000; 2000US-0218290.

26-JUL-2000; 2000US-0220963.

26-JUL-2000; 2000US-0220964.

14-AUG-2000; 2000US-0224518.

14-AUG-2000; 2000US-0224519.

14-AUG-2000; 2000US-0225213.

14-AUG-2000; 2000US-0225214.

14-AUG-2000; 2000US-0225266.

14-AUG-2000; 2000US-0225267.

14-AUG-2000; 2000US-0225268.

14-AUG-2000; 2000US-0225270.

14-AUG-2000; 2000US-0225447.

14-AUG-2000; 2000US-0225757.

14-AUG-2000; 2000US-0225758.

18-AUG-2000; 2000US-0225759.

22-AUG-2000; 2000US-0226681.

22-AUG-2000; 2000US-0226868.

22-AUG-2000; 2000US-0227182.

23-AUG-2000; 2000US-0227009.

30-AUG-2000; 2000US-0228924.

01-SEP-2000; 2000US-0229287.

01-SEP-2000; 2000US-0229343.

01-SEP-2000; 2000US-0229344.

01-SEP-2000; 2000US-0229345.

05-SEP-2000; 2000US-0229509.

05-SEP-2000; 2000US-0229513.

06-SEP-2000; 2000US-0230437.

06-SEP-2000; 2000US-0230438.

08-SEP-2000; 2000US-0231242.

08-SEP-2000; 2000US-0231243.

08-SEP-2000; 2000US-0231244.

08-SEP-2000; 2000US-0231413.

08-SEP-2000; 2000US-0231414.

08-SEP-2000; 2000US-0232080.

08-SEP-2000; 2000US-0232081.

12-SEP-2000; 2000US-0231968.

14-SEP-2000; 2000US-0232397.

14-SEP-2000; 2000US-0232398.

14-SEP-2000; 2000US-0232399.

14-SEP-2000; 2000US-0232400.

14-SEP-2000; 2000US-0232401.

14-SEP-2000; 2000US-0233063.

14-SEP-2000; 2000US-0233064.

14-SEP-2000; 2000US-0233065.

KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
XX Homo sapiens.
OS WO200155320-A2.
XX 02-AUG-2001.
PD 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465570/50.
XX DR N-PSDB; AAL01300.
XX PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition
XX Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a protein of the invention.
XX SQ Sequence 147 AA;

Alignment Scores:
Pred. No.: 1.14e-38 Length: 147
Score: 503.00 Matches: 97
Percent Similarity: 80.15% Conservative: 12
Best Local Similarity: 71.32% Mismatches: 27
Query Match: 12.55% Indels: 0
DB: 22 Gaps: 0

US-09-743-237-1 (1-2241) x AAM95330 (1-147)

Qy 495 GTTGAATCAAGAGAGCTGTGCTGCCAGGCGCAGCTTTC 554
Db 8 ValGluLeuLysGluAlaGlyThrThrSerAsnAsnProGluGluAlaThrLeu 27
Qy 555 CAGGCCCTCTGGCTCAGGAATCTGTGCAAGTTCCTCCATCATCCAGGAGGAGGAG 614
Db 28 GlnAsnLeuLeuAlaGlnGluSerCysCysLysPheProSerGlnGluLeuGluAsp 47
Qy 615 GCCTCCAGCTCCCTCGGAAGAAGACTCCAGCCCATGTTGCTCAGCTCAAGGA 674
Db 48 AlaSerCysSerLeuLysLysAspSerAsnProMetValIleCysGlnLeuLysGly 67
Qy 675 GGCCTCCAGATGCTCTCATAGACAACCTGTGGCCGAGGAGCTCAAGGCTCCATCTG 734
Db 68 GlyThrGlnMetLeuCysIleAspAsnSerArgThrArgGluLeuLysAlaLeuHisLeu 87
Qy 735 CTTCCTCAGTACGATGACGAGCAGTTCCTCAGTCAGAGCTCCCTAAGCCCAATGACA 794
Db 88 ValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSerAspValProLysProMetThr 107
Qy 795 ACTTTAGTGGGAAGACTTCTCCAGTACCAGCGAAGTAAATCTCATCACACAGGTTGAT 854
Db 108 AlaLeuValGlyArgPheLeuProAlaSerThrLysLeuAsnLeuIleThrGlnGlnLeu 127
Qy 855 AATGAGCTCTCCATCAGCTGTCAATGGGCTGCCTTTCCCTCTGGA 902
Db 128 GluGlyAlaLeuProSerValValAsnGlySerAlaPheProSerGly 143

RESULT 5
ABB62035
ID ABB62035 standard; Protein; 950 AA.
XX AC ABB62035;
XX DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 12897.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX OS Drosophila melanogaster.

XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL06138.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions
XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 950 AA;

Alignment Scores:
Pred. No.: 2.28e-32 Length: 950
Score: 440.50 Matches: 138
Percent Similarity: 45.87% Conservative: 62
Best Local Similarity: 31.65% Mismatches: 141
Query Match: 10.99% Indels: 95
DB: 22 Gaps: 14

US-09-743-237-1 (1-2241) x ABB62035 (1-950)

Qy 257 CCCTCTGTAGATAGGGAGAGTCTTTAAACCTATGTAGAGTTTGTGAATTCGTCT 316
Db 572 ProValGlnLysLeuIleAsnArgThrAlaAsnValGlnArgValValSerSerThr 591
Qy 317 TCTCAACCATATGCTAAGCTATATAGCAATTCCTTCAAAATTCGTATATACTAGGAGA 376
Db 592 SerPro-Ser-----SerAsnSerThrLysLysIleThrAsnThrValG 606
Qy 377 ACCTCTGATTCTCTCGCTCTACATCCTAGTCTAGTGTACAGGGGAAATCATTTTG 436
Db 606 nPro----- 607
Qy 437 GTGAGACTCCGATGAAGTACTGTCAGGTTCCCAAGGAGCAGCAAGCAAGAAAAGTGT 496
Db 608 -----ThrGlySerAsnGlyAlaLysTyrMetIleCysAsnSe 620
Qy 497 TGAATCAAAAGACAGGTGGTAGTGCCA-----GGCGGAGCCCTGAAGA 544
Db 620 rGlyValProGlnSerSerThrSerAlaMetArgArgGlyThrGlyThrValGluAs 640
Qy 545 CGCAGCTTTCCAGGCCCTCTGGCTCAGGAATCCTGTGCAAGTCCCATCATCCAGGA 604
Db 640 n-LysThrArgArgPro-----ProProIleSerProGlnGlnHisArgPheL 656
Qy 605 GGCAAGGAGGCTCCAGTGCCTCGGAAGAAGACTCCAGCCCATGCTGATTGTCA 664

Db 656 ysGlnMetGlyPro-GlnGlnSerLys-----HisGlnGlnLeuGln 670
 Qy 665 GCTGAAGAGGCGCCAGAGTCTGCATAGACAACAGTGTGGCGGAGGAGCTCAAGC 724
 Db 671 AlaGlnAlaLysGln-----ArgIleArgGlnGlnLeuProThrGluGlnSer 687
 Qy 725 GCTCATCTGCTTCTCAGTACGATGACGAGCAGTGTCCCTCAGTACGAGCTCCCTAA 784
 Db 688 ThrProIleLysValGluProLysLeuProThrLeuProGlyValLysAla-AsnVa 707
 Qy 785 GCCAATGACAACCTTAGTGGGAGACTTCTGCCAGTACACGGAAGTAAATCTCATCAC 844
 Db 707 lProAlaLysProLeuPheGluValLeuLysPro---ProAla----- 720
 Qy 845 ACAGGTTGATATGAGCTCTCCCATCAGCTCAATGGGCTGCTTCCCTCTGGACC 904
 Db 721 -----ThrAlaAlaAlaGlyAlaValAspProLeuGlyGln 733
 Qy 905 TGCTCTGCAAGGCGCCCAAAATAACTCTGCTGGG-----TACTG 946
 Db 733 yMetThrSerArgArgLysHisCysAsnCysSerLysSerGlnCysLeuLysLeuTyrCy 753
 Qy 947 TGACTGCTTCTCCACGGGAGCTTCTGCAACAGCTGCAGCTGC-----AACAA 994
 Db 753 sAspCysPheAlaAsnGlyGluPheCysGlnAspCysThrCysLysAspCysPheAsnAs 773
 Qy 995 CTGCGCCATGAGTCGAGCGCTTCAAGCCATAAGCGGTGCTGTGTAGTAAATCTCTGA 1054
 Db 773 nLeuAspTyrGluValGluArgGluArgAlaIleArgSerCysLeuAspArgAsnProSe 793
 Qy 1055 AGCTTTTCAACCAAAATGGGAAAGCGCTGCGAGCTGTAACTTCGACAGCA 1114
 Db 793 rAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetArgLeu---HisAsn 812
 Qy 1115 AGGTGCAACTGTAAAGCGCTCAGCGTCTGCTGAAGAACTACTGTAGTGTATGAGGCCAA 1174
 Db 812 sGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaL 832
 Qy 1175 AATCATGTGTTCTTCATTTGCAATGCTGCTTGCATAAACTATGAAGAAGTCCAGA 1234
 Db 832 sIleProCysSerSerIleCysLysValCysValCysArgAsnMetGluAspArgProAs 852
 Qy 1235 AGAAAAATGCTGTATGATGACACACCCCTACTACATGAGCGCTGGGGACTTTGAGAGCAGCCA 1294
 Db 852 p-----ValAspMetAspSerLeuAs 859
 Qy 1295 TTATTTGTCGCCAGCAAGTCTCTAGGACCTCCAAAACCTGAGAAAAATAGGACG----- 1349
 Db 859 pGlyLeuMetGlyValGluGlyGlnLysLysAspLys---AlaLysAsnLysGlnLeuAs 878
 Qy 1350 -----GCCTTCTCTGTATCTCTGGAAGTACTGAGGAGCCACATGTGCTGCTCT 1399
 Db 878 nGluAsnArgAlaAsnIleTyrPheThrAspAspValIleGluAlaThrIleMetCysWe 898
 Qy 1400 GCTGCCCGCAGGTGAGGAGCAGCAGCAGTGTGTTCCCAAGCTTGGCTGAGCAGT 1459
 Db 898 tIleSerArgIleValMetHisGluLysGlnAsnValAlaValGluAspMetGluArgGln 918
 Qy 1450 GATCTCGAGGAGTTGGAAGGTGCTGTCGAGATTCTC 1499
 Db 918 uValMetGluMetGlyGluSerLeuThrGlnIleIle 931
 RESULT 6
 AAW83392
 ID: AAW83392 standard; Protein: 438 AA.
 XX
 AC AAW83392;
 XX
 DT 29-MAR-1999 (first entry)
 XX
 DE Caenorhabditis elegans synMuv protein LIN-54.
 XX
 KW LIN-54; synthetic multivulvar; SynMuv; signal transduction;

KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
 cell proliferation; gene therapy.
 XX Caenorhabditis elegans.
 OS W09854299-Al.
 PN 03-DEC-1998.
 PD 28-MAY-1998; 98WO-US11043.
 XX 28-MAY-1997; 97US-0047996.
 PF (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PI Ceol C, Horvitz HR, Lu X;
 DR WPI; 1999-045362/04.
 DR N-PSDB; AAV72865.
 XX Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
 useful for treating diseases associated with altered levels of cell
 proliferation, e.g. carcinomas
 PS Claim 7; Fig 13; 70pp; English.
 XX This is the amino acid sequence of LIN-54, a novel protein of
 Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
 synthetic multivulvar (synMuv) gene involved in cell fate and cell
 proliferation, and is part of a pathway that may be used as a
 genetic and biochemical model system for tumour suppression and
 cancer in mammals. SynMuv pathway genes and proteins may be used
 to identify genes which are part of the mammalian pathway and
 to identify genes, proteins and therapeutic compounds which
 modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
 C. elegans SynMuv polypeptides selected from LIN-37, -35, -55, -52,
 -53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
 (1) nucleic acids (see AAV72866-67) encoding the mouse and human
 homologues of C. elegans LIN-54; (2) vectors containing the nucleic
 acids; (3) transgenic cells; (4) a pure mammalian SynMuv
 polypeptide, and (5) an antibody which binds to a SynMuv family
 protein. The SynMuv nucleic acids and polypeptides can be used to
 diagnose and treat, especially by gene therapy, conditions
 involving altered levels of cell proliferation, e.g. SynMuv-
 associated carcinomas.
 XX SQ Sequence 438 AA;
 Alignment Scores:
 Pred. No.: 3.25e-29 Length: 438
 Score: 405.50 Matches: 132
 Percent Similarity: 44.37% Conservative: 57
 Best Local Similarity: 30.99% Mismatches: 158
 Query Match: 10.12% Indels: 79
 DB: 20 Gaps: 18
 US-09-743-237-1 (1-2241) x AAW83392 (1-438)
 Qy 419 CAGGGGAAATCATTTTGTGAGACTCCGATGAACACTGTCAGGTCCCAAGCAGCAA 478
 Db 3 GlnGlyGluIleVal-TyrGlnAspAspAspTyrTyrAspGluSerGluIleTyrAs 22
 Qy 479 GCAACCAAGAAAAGTGTGAATCAAGACAGCGTGTAGTGTGCCAGGCGGAGCC 538
 Db 22 pAsnTyrGluGluGlyAlaGluPheIleGluValAsnGlyGlnLeuValProHisAsnPr 42
 Qy 539 TGAGACGACGCTTTCAGGCGCCCTCTGCTCAGGAATCTCTTCAAGTCCCATCATC 598
 Db 42 O-AsnLeuGlnAlaGlnGln-----AsnArgProGlyThrSerMetIleG 58
 Qy 599 CCAGGAGCAGAGGAGGCGCTCCAGCTGCCTCGGAAGAAAGACTCCAGCCCCATGAT 658
 Db 58 InGlnHisAsnArgSerMetGluValAsnGlnGlyLeuValLysAspGluProIle-Asp 77


```
|||||
Db 114 TyrCysAspCysPheAlaAsnGlyGluPheCysAsnAsnCysAsnCysThrAsnCysTyr 133
QY 990 AACAACTGCCCGCCATGAGCTCGAGCGCTTCAAAGCCATAAAGGGGTGCTTTGATAGAAAT 1049
Db 134 AsnAsnLeuGluHisGluAsnGluArgGlnLysAlaTleLysAlaCysLeuAspArgAsn 153
QY 1050 CCTGAAGCTTCCCAACCAAAATGGGAAAGCCGCTCGGGAGCTGCTAAACTTCGACAC 1109
Db 154 ProGluAlaPheLysProLysIleGlyLysGlyLysGluGlyGluSerAspArgHis 173
QY 1110 AGCAAGGTCGAACCTGTAACCGCTCAGGCTGCTGAAGAACTACTGTGAGTGTATGAG 1169
Db 174 SerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGlu 193
QY 1170 GCCAAATCATGCTGTTCTTCATTGCAATGCAATGCAATGCTTCCAAAACTATGAAGAAAT 1229
Db 194 AsnIlePheIleTyrPheCysCysArg-----TyrArgGlnIle 207
QY 1230 CCAGAACAAAATGCTGATGAGCAGACACCCCACTACATGAG 1271
Db 208 LeuGluLysAlaIleGlnLeuSerGlyAlaGluGlnLeuGlu 221
RESULT 8
AAM93348
ID AAM93348 standard; Protein: 251 AA.
XX AC
XX AAM93348;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2895.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94268.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
```

```
SQ Sequence 251 AA;
Alignment Scores:
Pred. No.: 1.07e-23 Length: 251
Score: 346.00 Matches: 66
Percent Similarity: 80.65% Conservative: 9
Best Local Similarity: 70.97% Mismatches: 18
Query Match: 8.63% Indels: 0
DB: 22 Gaps: 0
US-09-743-237-1 (1-2241) x AAM93348 (1-251)
QY 495 GTTCAATCAAGAGAGAGTGGTAGTGTCCAGCGCGCAGCCCTGAAGACGACGCTTTC 554
Db 158 ValGluIleLysGluAlaGlyClyThrThrSerAsnAsnProGluGluAlaThrLeu 177
QY 555 CAGCCCCCTCGGCTCAGGAATCCTGTTGCAAGTTCATCCATCCAGGAGGAGAGGAG 614
Db 178 GlnAsnLeuLeuAlaGlnLysCysCysLysPheProSerSerGlnGluLeuGluAsp 197
QY 615 GCCTCCAGCTGCCCTCGGAGAACAAAGACTCCAGCCCATGTTGTCAGCTGAAGGA 674
Db 198 AlaSerCysCysSerLeuLysLysAspSerAsnPrometValIleCysGlnLeuLysGly 217
QY 675 GCGCCAGAGTCTGTCATAGACAACTGTGGCGGAGGAGCTCAAGAGCCTCCATCTG 734
Db 218 GlyThrGlnMetLeuCysIleAspAsnSerArgThrArgGluLeuLysAlaLeuHisLeu 237
QY 735 CTTCCTCAGTACGATGACCAAGCAGAGCAGTTCCTCCCTCAGTCA 773
Db 238 ValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSer 250
RESULT 9
AAG76158
ID AAG76158 standard; Protein: 53 AA.
XX AC
XX AAG76158;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5922.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 11.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH35563.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 8365; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC
```

CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB7789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 53 AA;

Alignment Scores:

Pred. No.:	9.71e-15	Length:	53
Score:	248.00	Matches:	48
Percent Similarity:	94.34%	Conservative:	2
Best Local Similarity:	90.57%	Mismatches:	3
Query Match:	6.19%	Indels:	0
DB:	22	Gaps:	0

US-09-743-237-1 (1-2241) x AAG76158 (1-53)

QY 1377 GTGGAGGCCACATGTGCTGCTGCTGCGCCAGGCTGAGGAGCAGGAGGACTCT 1436
Db 1 ValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluAlaGluLysGluHisCys 20
QY 1437 TCCCAAGCTTGCTGAGCAGATGATCCCTGGAGGAGTTTGGAGGTCCTGTGCAGATT 1496
Db 21 SerLysCysLeuAlaGluGlnMetIleLeuGluGluPheGlyArgCysLeuSerGlnIle 40
QY 1497 CTCACATGATGTTCAAGTCCAGGCGCTGGAATTTAG 1535
Db 41 LeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 53

RESULT 10

ID ABB68888 standard; Protein; 243 AA.

XX
AC ABB68888;

XX
DT 26-MAR-2002 (first entry)

XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33456.

XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX
OS Drosophila melanogaster.

XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.

XX
PF 23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.

XX
PR 11-JUL-2000; 2000US-0614150.

XX
PA (PEXE) PE CORP NY.

XX
PI Venter JC, Adams M, Li PWD, Myers EW;

XX
DR WPI; 2001-656860/75.

XX
DR N-PSDB; ABL12991.

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions

XX

PS Disclosure; SEQ ID NO 33456; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 243 AA;

Alignment Scores:

Pred. No.:	1.48e-08	Length:	243
Score:	186.00	Matches:	59
Percent Similarity:	36.64%	Conservative:	26
Best Local Similarity:	25.43%	Mismatches:	61
Query Match:	4.64%	Indels:	86
DB:	22	Gaps:	8

US-09-743-237-1 (1-2241) x ABB68888 (1-243)

QY 1044 AGAATCTCTGACGTTTCCAAACCAAAATGGGGAAGCCGCTGGAGCTGCTAAACTT 1103
Db 7 ArgSerValAspLysAlaAspGlyLysGlyGln---GlyAlaGlyVal 25

QY 1104 CGACACAGCAAGGGTGCAACTGTAGCGCTCAGGCTGCTGAAGAACTACTGTGAGTGC 1163
Db 26 -----LysGlyCysCysCysLysArgSerGlnCysIleLysAsnThrCysAspCys 42

QY 1164 TAGAGGCCAAATCATGTGTTCTTCATTTGCAATTCGAAATTCGCAAAATCATGAA 1223
Db 43 TyrGlnSerMetAlaIleCysThrLysPheCysArgCysValGlyCysArgAsnThrGlu 62

QY 1224 -----GAAAGTCCAGAACCA--- 1238
Db 63 ValArgGluLeuValAspProAsnSerValAlaLysAsnSerSerAlaValLysArgGln 82

QY 1238 ----- 1238

Db 83 LysAlaAlaAlaMetSerAlaLysAlaAlaAlaAlaAlaLysAlaGlyIleAspVal 102

QY 1239 -----AAAATGCTGATG 1250
Db 103 GlnGlyLysAlaLeuGlnValAlaAlaSerThrLeuAlaLeuProGlyLysAlaLeuMet 122

QY 1251 AGCACACCCCTACATG-----GAGCCTGGGAGCTTTCAGAGAGCCCATTTAT 1298
Db 123 ThrProProLysThrLeuValAlaGlyLysPro---ProMetAlaSerSerHis--- 140

QY 1299 TTGTCCCCAGCCCAAGTTCTCAGGACCT----- 1325
Db 141 IleAsnProIleProIleSerArgProIleAlaThrAlaAlaThrProAlaArgAlaVal 160

QY 1326 -----CCAAACTGAGAAAAAT 1343
Db 161 LysGlnProAlaGluProProMetProValAsnLeuIleIleProValArgHisAspAsp 180

QY 1344 AGGCAGGCTTCTCCTGTATCTCCTGGGAAGTAGTGGAGGCCACATGTGCCCTGCTGCTG 1403
Db 181 ArgArgAspArgAsnLeuPheValGlnProValAsnAlaAlaLeuLeuGluCysMetLeu 200

QY 1404 GCCCAGGTGAGGAAGCAGCAGCAGGAGCAGCTGTTCCCAAGCTTGGCTGAGCAGATGATC 1463
Db 201 IleGlnAlaThrGluAlaGluGlnLeuGlyLeuAsnGluLeuGlnValCysGlnLeuVal 220

QY 1464 CTGGAGGAGTTTGGAGGTGCTGTCGAGATTC 1499
Db 1464 CTGGAGGAGTTTGGAGGTGCTGTCGAGATTC 1499

Db 2550 ----- LeuGluAla----- 2552
 QY 747 GATGACGAGCAGTTTCCTCAGTCAGCAGCTCCCTAAGCCAAATGACAAC-----TTA 800
 Db 2553 GlnAspGlnGlyValAlaProValSerProGlySerSerProThrAlaAlaProGlyLeu 2572
 QY 801 GTGGGAAGACTTCTGCCA-----GTACACGGAAGTTAAATCTCATCACACAG----- 848
 Db 2573 ValArgArgValSerProArgGlyHisProGlyProLeuGlyAlaSerValGlnProPro 2592
 QY 849 -----GTTGATATGAGCTCTCCCATCAGCTGCTCAATGGGCTGCC 890
 Db 2593 ValAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAla-----GlyGlyTrp 2610
 QY 891 TTTCCCTCTGGACCT----- 905
 Db 2611 GlyProTrpGlyProTrpSerHisCysSerArgSerCysGlyGlyClyLeuArgSerArg 2630
 QY 906 -----GCTCTGCAAGGGCCACCCAAATAACTCTGTCTGGTACTGTGAC---TGCTTC 956
 Db 2631 ThrArgAlaCysAspGlnProProGlnGlyLeuGlyAspTyrCysGlnGlyProArg 2650
 QY 957 TCCAGGGGGACTCTGCAACAGC----- 980
 Db 2651 AlaGlnGlyGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGlu 2670
 QY 981 -----TGC-----AGCTGCAACAACCTGGCCAT 1004
 Db 2671 GlyAlaGluTyrSerProCysGlyProProCysProArgSerCysAspLeuValHis 2690
 QY 1005 GAGCTCGAGCGTTCAAGCCATAAAGCGTGTCTTGATAGA---AATCCTGAAGCTTTC 1061
 Db 2691 -----CysValTipArgCysGlnProGlyCysTyr 2700
 QY 1062 CAACCAAAA----- 2720
 Db 2701 CysProProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnProGlyHisCys 2720
 QY 1089 GGAGCTGTCTAAACTT-----CGACACAGCAAGAGGTGCAACTGTAAGCGC--- 1133
 Db 2721 SerCysLeuAspLeuThrGlyGlnArgHisHisProGlyAlaArgLeuAlaArgPro 2740
 QY 1134 TCAGCTGCTGGAAGAACTACTGAGTCTATGAGGCCAAATAATGATGTCTTCTCCAPT 1193
 Db 2741 AspGlyCys-----AsnHisCysThrCysLeuGluGlyArgLeuAsnCysThrAspLeu 2758
 QY 1194 TGCATATGATTCGTTGCATAAACTATGAAGAAATGCCAAGACGAAATGCTGATGAGC 1253
 Db 2759 -----ProCysProAspCys----- 2763
 QY 1254 ACACCCCACTACATGGAGCCTGGGACTTTGAGAGCAGCATTATTGTCCCGCAGCAG 1313
 Db 2764 -----GlyGlyGlyGlnSerLeuHis-ProCysGlyGlnProCys 2776
 QY 1314 TTCTCAGGACCTCCAAACTGAGAAAAATAGCAGCGCTTCTCTGTATCTCTGGGAA 1373
 Db 2776 sProArgSerCysGlnAsp-----LeuSerProGlySe 2787
 QY 1374 GTAGTGGAGCCACATGTCCTGCTGCTGGCCAGGGTGAGAGCAGCAGGAGAC 1433
 Db 2787 rValCysGlnProGlySerValClyCysGlnPro-----ThrCysGln 2801
 QY 1434 TGTTCCTCCCAAGCTTGCTCAGCAGATGATCCTGGAGAGTTTGGAGGTGCTGCGCAG 1493
 Db 2801 yCysProLeuGlyGlnLeuSerGln-----AspGlyLeuCys----- 2813
 QY 1494 ATTCTCCACATGAGTTCAAGTCCAAGGGCTGAAAATTGAGTAGCGTGAAGCTGTAA 1553
 Db 2814 -----ValProProAlaHis 2818
 QY 1554 AGGGGATCCCTGTGGCAAGCCCTCAGCCCTGGGAATCTCCACCGAGGAAGCTGGTCCCA 1613
 Db 2818 sCysArgCysGlnTyrGlnProGlyAlaMetGlyIleProGluAsnGlnSerArgSerAl 2838

QY 1614 GGGAGGAGCAGAGCGCGCATCATG-----CCAGGTGAGCTGTGAGGTCTG 1661
 Db 2838 agly-SerArg---PheSerSerTrpGluSerLeuGluProGlyGluVal----- 2853
 QY 1662 AGTGATCTGCATGGTACTGGC-----CAGCCTACTCAAGGTATCTCTAAAG 1706
 Db 2854 -----ValThrGlyProCysAspAsnCysThrCysValAlaGlyIleLeuGln 2869
 QY 1707 TCAAGCAGCGCA-----GAGCCA-----CCCTGGGGA---TGGACA 1739
 Db 2870 CysGlnGluValProAspCysProAspProGlyValTrpSerSerTrpGlyProTrpGlu 2889
 QY 1740 CTGGCCCTCTCTCTGGGAGCGCTCTGGGACTCCCTG----- 1781
 Db 2890 AspCysSerValSerCysGly-----GlyGlyGluGlnLeuArgSerArgCysAla 2907
 QY 1782 -----CCCTGCATAAAAAGAGGGTATTTCTACTTGTGTATGTGTCTTCAAA 1835
 Db 2908 ArgProProCysProGlyProAlaArgGlnSerArgThrCys----- 2921
 QY 1836 TTGCTTAGTAGTACCTCCATTCAAGTTATATGAGCCAGCCTCAAGTTAGAGAGTAGGC 1895
 Db 2922 -----SerThrGlnVal-----CysArgGluAlaGly 2930
 QY 1896 TCTTCTTCAAGTGGACTC-----TGCCCAATACATACAGTCAGGTGGCCATCA 1946
 Db 2931 CysProAlaGlyArgLeuTyrArgGluCysGln----- 2941
 QY 1947 GGGGTTTTCCAGCGCAGCCTGTGACAGAGATATGGAGGGGCGGTAGAGCTG 2006
 Db 2942 -----ProGlyGluGlyCysProPheSerCysAlaHis---ValThrGlnGlnVal 2957
 QY 2007 GGTGTTGTTGGATTTTTCGGTTTTTCTCTCTGATTTCTGCTGAAGTGAAGAACT 2066
 Db 2958 Gly-----CysPheSerGluGlyCysGluGly 2967
 QY 2067 TCTCTCTCTGCTCA-----ACCTTTTCTCCATAATTACTGTGCACGGTGCCTGCTGAC 2120
 Db 2968 Cys-HiscysProGluGlyThrPheGln-HisArgLeuAlaCysValGlnGluCys---P 2986
 QY 2121 CAGTCACAGTGCACCTCA 2137
 Db 2986 roCysValLeuThrAla 2991
 RESULT 13
 AAU80189
 ID AAU80189 standard; Protein; 966 AA.
 XX
 AC AAU80189;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human TSPl domain containing protein encoded by cDNA FGI1869.
 XX
 KW TSPl; thrombospondin domain; FGI1869; angiogenesis; vasculogenesis;
 KW human.
 XX
 OS Homo sapiens.
 XX
 PN JP2002085059-A.
 XX
 PD 26-MAR-2002.
 XX
 PF 08-SEP-2000; 2000JP-0273778.
 XX
 PR 08-SEP-2000; 2000JP-0273778.
 XX
 PA (KAZU-) 2H KAZUSA DNA KENKYUSHO.
 PA (YOSH) YOSHITOMI PHARM IND KK.
 XX
 DR WPI; 2002-378268/41.
 DR N-PSDB; ABK50390.

XX TSPI domain-containing polypeptide useful for drug compositions -
 XX Claim 1; Page 30-33; S1pp; Japanese.
 XX The invention relates to a TSPI (thrombospondin 1) domain-containing
 CC polypeptide comprising the proteins appearing as AAU80188 and AAU80189,
 CC encoded by cDNAs designated FG06969 and FG01896. Also included are
 CC proteins that are 50% homologous to the proteins and a polypeptide having
 CC at least one deletion, replacement, addition or insertion of amino acid
 CC in the proteins and having at least 8 repetitions of the TSPI domain.
 CC The polypeptide can be used in drug compositions particularly
 CC for disorders associated with angiogenesis and vasculogenesis. The
 CC present sequence is the TSPI domain containing protein encoded by
 CC cDNA FG01896.

XX SQ Sequence 966 AA;

Alignment Scores:
 Pred. No.: 0.00138 Length: 966
 Score: 136.50 Matches: 120
 Percent Similarity: 29.83% Conservative: 53
 Best Local Similarity: 20.69% Mismatches: 204
 Query Match: 3.41% Indels: 203
 DB: 23 Gaps: 29

US-09-743-237-1 (1-2241) x AAU80189 (1-966)

QY 435 TGGTGAGACTCCCGATGACTGCTCCAGGTTCCTCCAGGAGCAAGCAAGCAAGCAAGCAAGT 494
 DB 237 TTPSerTrpCysAspArgSerCysGlyGlyGlnSerLeuArgSerArgSerCysSer 256
 QY 495 GTTCAATCAAGAGCAGGTGTGTGCTCCAGGCGCCGCTGAGACGCGACTTC 554
 DB 257 SerProSerProLysAsnGlyGlyAlaProCysAlaGlyGluArgHisGlnAlaArgLeu 276
 QY 555 CAGGCCCTCTGGCTCAGGATCTGTGCAAGTTCATCCATCCAGGAGGAGGAGGAG 614
 DB 277 CysAsnProMetProCysGluAlaGlyCys-----ProAlaGlyMetGluValValThr 294
 QY 615 ---CCCTCCAGCTCCCTCCGAG---AAAGACTCCAGCCCGCCATGGTGATTTGTCACGTG 668
 DB 295 CysAlaAsnArgCysProArgArgCysSerAspLeuGlnGlyIleValCysGln--- 313
 QY 669 AAGGAGGCGCCAGATGCTCTGCTAGACAACTGTGGCGGAGGAGCTCAAAGCGCTC 728
 DB 313 ----- 313
 QY 729 CATCTGCTTCTCAGTACGATGACAGAGAGTTCCTCAGTCAGAGCTCCCTAAGCCA 788
 DB 314 -----AspAspGlnValCysGlnLysGlyCysArgCysProLys--- 326
 QY 789 ATGACAACTTTAGTGGGAAGACTTCTCCAGTACCAGGAGTAAATCTCATCACACAG 848
 DB 327 -----GlySerLeu-----Glu 330
 QY 849 GTTGATAATGAGCTCTCCATCAGCTGCTCAATGGGCTGCCCTTCCTCTGGACCTGCT 908
 DB 331 GlnAspGlyGlyCysValPro----- 337
 QY 909 CTGAAGGGCCACCAAAATAACTCTGTCTGGGTACTGTGACTGCTTCTCCAGC----- 962
 DB 338 -----IleGlyHisCysAspCysThrAspAlaGlnGly 348
 QY 963 -----GGGACTTCTGCAACAGCTCCAGCTGCTC----- 989
 DB 349 HisSerTrpAlaProGlySerGlnHisGlnAspAlaCysAsnAnCysSerCysGlnAla 368
 QY 990 AACAACTGCGCCATGAGCTCGAGCGCTTCAAAGCCCAATAAGCGCTCTTGATAGAAT 1049
 DB 369 GlyGlnLeuSerCysThrAlaGlnProCysProProThrHisCysAlaTrpSerHis 388
 QY 1050 CCTGAAGCTTTTCCAACCA---AAAATGGGAAAGCGCTCTGGGAGCTGCTAAACTTCGA 1106

DB 389 TrpSerAlaTrpSerProCysSerHisSerCysGlyProArgGly-----Gln 404
 QY 1107 CACAGCAAGGGTGCACAACTCTAAGCGCTCAGGC-----TGCCTGAAGAAGTACTGTGAG 1160
 DB 405 GlnSerArgPheArgSerCysThrProGluGlyValIleCysGluAspThrGluCysAla 424
 QY 1161 TGCTATGAGCCCAAAATCATGTCTTCCATTTGCAAAATGC---ATTGCTTGCAAAAC 1217
 DB 425 ValProGluAlaTrpThrLeuTrpSerSerTrpSerAspCysProValSerCysGlyGly 444
 QY 1218 TATGAAGAAAGTCCAGAACGAAATAATCTGATGACACACCCACCTATGGAGCGCTGG 1277
 DB 445 GlyAsnGlnValArgThrArgAlaCysArgAlaAlaProHisHisArgSerPro--- 463
 QY 1278 GACTTTGAGAGCAGCCATTTATTTGCTCCCAAGCAAGTTCTCAGGACCTCCAAACTGAGA 1337
 DB 464 -----ProCysLeuGlyProAspThrGlnThr 472
 QY 1338 AAAAATAGCAGGCC-----TTCTCTGTATCTCTCGGAAGTAGTGGAGGCCACA 1388
 DB 473 ArgGlnGlnProCysProGlyLeuLeuGluAlaCysSerTrpGlyPro-TripGlyProCys 492
 QY 1389 TGTGCTCTGCTGCTGCGCCAGGTGAGAAAGCAGACAGCAGGAGCAGCTGTCCCAAGCTTG 1448
 DB 492 sSerArgSerCys-GlyProGlyLeuAlaSerArgSerGlySer-Cys---ProCysLeu 510
 QY 1449 GCTGACGAGATGATCTCTGGAGGAGTTGGAAAGTGTGCTGTGCGAGATTTCTCCACATCGAG 1508
 DB 511 -----MetalAlaLysAlaAspProThrCysAsnSerThrPheLeuHisLeuAsp 526
 QY 1509 TTCAAGTCCAAAGGGCTGAAATTTAGTAGCTGCAAGCTGGTAAAGGG----- 1557
 DB 527 Thr-----GlnGlyCysThr-SerGlyProCysProGly 537
 QY 1558 -GAATGCTCTGCAAGCC---TCAGCCCTGGGAATCTGCACCGAGGAAGCTGTGCGCCA 1613
 DB 537 uGluCysValTrpSerSerTrpSerTrpThrArgCysSer----- 551
 QY 1614 GGGAGGAGCAGAGGCGCGCATCATGCCAGGTGAGTGTGAGTGTGAGTGTGATCTGCAT 1673
 DB 552 -----CysArg 553
 QY 1674 GGTACTGGCCGCGCTACTCAAGGTATCTAAAGTGAAGCAGGAGCAGCCCTGGGGA 1733
 DB 553 gValLeuValGln-----GlnArgTrpArgHisGlnGly-- 564
 QY 1734 TGGACTGCGCCCTCTGCTGCTGGGAGGCGCTCTGGGAGCTCCCTGCGCTGCATFAAA 1793
 DB 565 -----ProAlaSerArgGlyAlaArgAlaGlyAlaProCysThrArg---Le 579
 QY 1794 AGAGGTGATTTTCTACTTCTTGTATGTCTTGTCTTCAAAATTC-----TTAGT 1844
 DB 579 uAspGlyHisPheArg-----ProCysLeuIleSerAsnCysSerGluAspSerCys 596
 QY 1845 AGTACTCTCAATCAAGTTATTATGAGCCAGCCTCAAGTTAGAGAGCTAGCTCTTCTTCA 1904
 DB 596 sThrProProPheGlu-PheHis-Ala-----CysGlySerProCys 609
 QY 1905 GGTGGACTTCCCAATACATACATAA----- 1931
 DB 610 AlaGlyLeuCysAlaThrHisLeuSerHisGlnLeuCysGlnAspLeuProProCysGln 629
 QY 1932 -----GTCAAGTGGCCATCAGGGGTTTTCAGGCCAG 1964
 DB 630 ProGlyCysTrpCysProLysGlyLeuLeuGluGlnAlaGlyGlyCysIleProProGlu 649
 QY 1965 GCCTGTGAC-----AGGAGATATGGAGGGGCTCGGGTTAGAGCTG 2006
 DB 650 GlucysAsnCysTrpHisThrSerAlaAlaGlyAlaGlyMetThrLeu 665

RESULT 14
 AAW87504


```
Db 476 hrSerAlaArgTrpArgThrArgSerProAlaSerArgTrpAlaLeuArgLeuSerLeuS 496
Qy 708 CGCCACAGTTCTCTATCAGAGCATCTGGCGCTCTTTCAGCTGACAAATCACCATGG 649
Db 496 erProGlySerTrpMetProSer-SerMetMetLeuLeuSerSer-----ThrThrTrp 513
Qy 648 GCCTGGAGTCTTTCTCCGAGGGCAGCTGGAGGCT-----CCCTCTG 607
Db 514 --GlnAlaArgThrArgAlaAlaSerTrpSerProLeuGlyLeuAlaArgSerLeuLeu 532
Qy 606 CCTCCTGGGATGATGGGAACCTGCAACAGGATTCCTGAGCAGAGGGCCCT----- 556
Db 533 ProLeuAlaThrAlaSerProCysArgArgThrProThrGlySerGlyProThrTrpArg 552
Qy 555 -----CysCysTrpTrpProTrpGlyTrpProCysTrpSerSerProG 609
Qy 459 GCAGTACTTCATCG 446
Db 609 lySerThrTrpSer 613

RESULT 15
AAU80188
ID AAU80188 standard; Protein; 961 AA.
XX
AC AAU80188;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human TSPl domain containing protein encoded by cDNA FGO6969.
XX
KW TSPl; thrombospondin domain; FGO6969; angiogenesis; vasculogenesis;
KW human.
XX
OS Homo sapiens.
XX
PN JP2002085059-A.
XX
PD 26-MAR-2002.
XX
PF 08-SEP-2000; 2000JP-0273778.
XX
PR 08-SEP-2000; 2000JP-0273778.
XX
PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
PA (YOSH ) YOSHITOMI PHARM IND KK.
XX
DR WPI: 2002-378268/41.
DR N-PSDB; ABK50389.
XX
TSPl domain-containing polypeptide useful for drug compositions -
XX
Claim 1; Page 23-26; 51pp; Japanese.
XX
CC The invention relates to a TSPl (thrombospondin 1) domain-containing
CC polypeptide comprising the proteins appearing as AAU80188 and AAU80189,
CC encoded by cDNAs designated FGO6969 and FGO1896. Also included are
CC proteins that are 50% homologous to the proteins and a polypeptide having
CC at least one deletion, replacement, addition or insertion of amino acid
CC in the proteins and having at least 8 repetitions of the TSPl domain.
CC The polypeptide can be used in drug compositions particularly
CC for disorders associated with angiogenesis and vasculogenesis. The
CC present sequence is the TSPl domain containing protein encoded by
```

```
CC cDNA FGO6969.
XX
SQ Sequence 961 AA;
Alignment Scores:
Pred. No.: 0.0109 Length: 961
Score: 127.00 Matches: 146
Percent Similarity: 30.27% Conservative: 58
Best Local Similarity: 21.66% Mismatches: 206
Query Match: 3.14% Indels: 264
DB: 23 Gaps: 42

US-09-743-237-1 (1-2241) x AAU80188 (1-961)
Qy 2159 AGCCACCTACCTTCTGGTCTCTGAGGTCTGCTGAGGTGCTGAGCAGCGGACCGTGCAGC 2100
Db 387 SerHisTrpSerAlaTrp-----SerProCysSer 396
Qy 2099 AGTAATTATGGAGAAAGTTTGGACAGGAGACAAAGTTTTCACCTTCAAGCAGAAATACA 2040
Db 397 HisSerCysGlyProArg--GlyGlnGlnSerArgPheArgSerSerThrSerGlySer 415
Qy 2039 GGAGAAAAA---CGCAAAAAATCCAAACAAACCCAGCTCTAACCCGACCCC----- 1989
Db 416 TrpAlaProGluCysArgGluGluGlnSerGlnSerGlnProCysProGlnProSerCys 435
Qy 1988 CCTCCCATTA-----TCTCTGTACAGGCGCTGG---CCTGG 1956
Db 436 ProProLeuCysLeuGlnGlyThrArgSerArgThr-LeuGlyAspSerTrpLeuGlnG 455
Qy 1955 AAAA-----ACCCCTGATGGCCACCTGACTTGTATGTATGTATGGGCA 1914
Db 455 yGluCysGlnArgCysSerCysThrProGluGly---ValileCys----- 469
Qy 1913 GAGTCCACCTGAGACAGAGCCTAGCTCTCTAAGTGGTGGCTCATATACTTGAAT 1854
Db 470 -----GluAspThrGluCysAlaValProGluAlaTrpThrLeuTrpSer-SerT 486
Qy 1853 GGAGGTACTACTAAGCAATTTGAAAGCAACACACATAACAAGTACAAATCACCCCTCT 1794
Db 486 rp-----SerAspCysProValSerCysGlyGlyGlyAsnGlnValArgThrA 502
Qy 1793 TTTTATGACAGGAGGAGTCCCGAGAGGCCCTCCCGAGGACAGGAGGCCAGTGTCCA 1734
Db 502 rgAlaCysArgAlaAlaAlaProHisArgSerProProCysLeuGlyProAspThrG 522
Qy 1733 TCCCCAGGTG-----CCTCTGCTGCTTGCACCTTTAGGATACCTTTCAGT 1689
Db 522 lnThrArgGlnGlnProCysProGlyLeuLeuGluAlaCysSerTrp----- 537
Qy 1688 AGGCTGGCCAGTACCATGCAGATCCTCAGACCTCAGACCTGACCTGCCATGATGCGCG 1629
Db 538 -----GlyProTrpGlyProCysSerArgSerCysGlyProGlyLeuAlaSerArgSerG 556
Qy 1628 GCCTCTGCTCTCCCTG-----GGCACCAGCTTCTCTCGGTG 1593
Db 556 lySerCysProCysLeuMetAlaLysAlaAspProThrCysAsnSerThrPheLeuHis 576
Qy 1592 CAGATTCACAGGCTGAGGCTTCCACAGCATTCCTCCCTTTACCAGCTTGCACGCTACTC 1533
Db 576 euAspThrGlnGly-----C 581
Qy 1532 AATTTTCAGCCCTTGGACTTGAACATGATGGAGAAATCGCAGACGACCTTCCAAA 1473
Db 581 ystyrSerGlyPro-----CysProGluAspSerCys-----T 592
Qy 1472 CTCTCTCA-----GGATCATCTGCTCAGCCAAAGCTTGGGAACAGTG 1431
Db 592 hrProPheGluPheHisAlaCysGlySer----- 602
Qy 1430 CTCTGCTGCTTCTCTCACCCTGGCCAGCAGGACGACATG----- 1387
Db 603 -----ProCysAlaGlyIleCysAlaThrHisLeuSerHisGlnLeuCysG 618
```

QY 588 --ACTTTGCAACAGGATTCTCTGAGCCAGAGGGGCTGGA 553
 Db 921 isLeuCysGlnArgValProGlyValProThrProGly 933

Search completed: April 21, 2003, 11:28:01
 Job time : 136.879 secs

QY 1386 -----TGGCTCCACTACTTCCAGGAGATACAGAGAGGCGCTGCTATTTTCTCAG 1332
 Db 618 lnAspLeuProProCysGlnProGlyCysTyrCysProlys-----GlyL 633
 QY 1331 TTTTGGAGTCTCTGAGAACTTGGCTGGGACAAATAATGGCTGCTCTCAAAAGTCCACAGG 1272
 Db 633 euLeuGlu-----GlnAlaGlyGlyCysIleProProGluGluC 646
 QY 1271 CTCCATGT---AGTGGGTGTGCTCATCAGCATTTTTCGTTCTG-----GACTTTC 1224
 Db 646 ysAsnCysTrpHisThrSerAlaAlaGlyAlaGlyMetThrLeuAlaProGlyAspArgL 666
 QY 1223 TTCATAGTTTTCACAAAGCATTTGCAAAATGGAAGACACATGATTTTGGCCTCATA 1164
 Db 666 euGlnLeuGlyCysLysGluCysArgArgGlyGluLeu-----HisC 682
 QY 1163 GCACCTACAGTAGTTCTTCAGGC-----AGCCTGAGCGCTTACAGT-----T 1122
 Db 682 ysThrSerGlnGlyCysGlnGlyLeuLeuProLeuSerGluTrpSerGluTrpSerProc 702
 QY 1121 GCACCCCTTTC-----TGCTGCGAAAGTTT 1098
 Db 702 ysGlyProCysLeuProProSerAlaLeuAlaProAlaSerArgThrAlaLeuGluGluH 722
 QY 1097 AGCAGCTC-----CCAGAGCGGCTTTCCTCCATTTTGGTTGAAAGCTTCAGGATTTCT 1044
 Db 722 lsTrpLeuArgaspProThrGlyLeuSerProThrLeuAlaProLeuLeuAlaSerGluG 742
 QY 1043 ATCAAGACAGCGCTTTATGGCTTTGAAGCGCTCGAGCTCAT-----GGCGCA---- 997
 Db 742 lnHisArgHisArgLeuCysLeuAspProAlaThrGlyArgProTrpThrGlyAlaProH 762
 QY 996 -----GGTTGTTGCAGCT 984
 Db 762 isLeuCysThrAlaProLeuSerGlnGlnArgLeuCysProAspProGlyAlaCysProA 782
 QY 983 GCAGCTGT-----TCCAGAAGTCCC----- 964
 Db 782 spSerCysGlnTrpSerLeuTrpGlyProTrpSerProCysGlnValProCysSerGlyG 802
 QY 963 -----CGCTGGAGACAGTCACAGTACC-----CAGACAGAGTTATTTT 924
 Db 802 lyPheArgLeuArgTrpArgGluAlaGluAlaLeuCysGlyGlyCysArgGluProT 822
 QY 923 GGCTGGCCCTTCAGAGAGTCCAGAGGGAAGGAGCCCATTTGACAGCTCATGGGAG 864
 Db 822 rp---AlaGlnaspArgLysLeuGlnArgAlaLeuProArgLeuGluLeuArgGlyP 841
 QY 863 AGCTCCATTATCAACCTGTGTGATGAGATTTAACTTCGCTGCTGCTGCGAGAGTCTTCC 804
 Db 841 roArgHis----- 843
 QY 803 CACTAAAGTTGCTATTGGCTTAGGAGCTCTGACTAGGGAAACTGCTGTGGTCATCGTA 744
 Db 844 -----CysIleHisProGlyLeuCysGlnProValProT 855
 QY 743 CTGAGGAAGCAGATGGAGCGCTTTGAGCTCCCTCGGCCACACATTTGCTCTATGCGAGCAT 684
 Db 855 hrGln-----LeuCysArgProL 861
 QY 683 CTGGG---CGCCTCCTTTTCAGCTGACAAA-----TCACCATGG- 649
 Db 861 euGlyProArgSerValSerAlaGlyThrLeuProProArgLeuProLeuSerProTrpP 881
 QY 648 -----GGCTGGAGTCTTTCTTCGAG-----GGCAGCTGGAGGCCT---- 613
 Db 881 roAlaGlyProGlyTrpAlaLeuCysAlaAspLeuLeuLeuProLeuTrpProProGlnC 901
 QY 612 -----CCTCTGCCTCTGGGATGATGGGA----- 589
 Db 901 ysGlnCysLeuLeuGlyAlaGlyProGlyProGlyAlaAlaGlyLeuProLysLeuH 921

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:21:28 ; Search time 23.4917 Seconds
(without alignments)
5613.620 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 Tatctctgtgggttgcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=exp
-Q/cgn2_1/USPTO.spool/US09743237/runat_21042003_111945_440/app_query.fasta_1.4494
-DB-issued_Patents_AA -QFWT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237@cgn2_1.66 @runat_21042003_111945_440 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	119.5	3.0	1185	4	US-09-041-886-23
C 2	119.5	3.0	2594	4	US-08-718-388-7
C 3	119.5	3.0	5405	4	US-08-718-388-9
C 4	118.5	3.0	3118	4	US-09-579-181-1
C 5	117.5	2.9	2972	4	US-09-579-181-2
C 6	113	2.8	498	2	US-08-660-963-12
C 7	112.5	2.8	323	1	US-07-913-107-2
C 8	112.5	2.8	323	1	US-08-459-201-2
C 9	112.5	2.8	323	1	US-08-281-248-2
C 10	112.5	2.8	323	5	PCR-US93-06421-1
C 11	112	2.8	652	2	US-08-751-305-2
C 12	110	2.7	830	6	5378464
C 13	108.5	2.7	610	3	US-08-365-470-3
C 14	108.5	2.7	610	3	US-09-209-668-19
C 15	108.5	2.7	610	4	US-09-009-490A-89
C 16	108.5	2.7	610	6	521870-2
C 17	107	2.6	1142	3	US-08-993-118-7
C 18	107	2.6	1142	3	US-08-845-528C-7
C 19	107	2.6	1142	4	US-09-061-709-2
C 20	106	2.6	1564	4	US-09-467-997-1
C 21	104	2.6	1093	3	US-08-545-860D-55
C 22	104	2.6	1093	5	PCR-US94-04496-55
C 23	101	2.5	1251	5	PCR-US95-02251-3
C 24	101	2.5	1252	1	US-08-199-780-3
C 25	101	2.5	1252	2	US-08-316-650-3
C 26	100.5	2.5	1253	3	US-08-479-722B-4
C 27	99	2.4	503	4	US-08-999-774A-10
C 28	98.5	2.4	2414	1	US-08-227-536-2
C 29	98.5	2.5	2414	5	PCR-US95-04682-2
C 30	97.5	2.4	2910	1	US-08-466-033-183
C 31	97.5	2.4	2910	2	US-08-444-733-183
C 32	97.5	2.4	2910	2	US-08-464-134-183
C 33	97.5	2.4	2910	2	US-08-461-361-183
C 34	97.5	2.4	2910	2	US-08-485-910-183
C 35	97.5	2.4	2910	5	PCR-US95-06266-157
C 36	96.5	2.4	426	1	US-07-918-023-2
C 37	96.5	2.4	689	4	US-09-177-249-2
C 38	96.5	2.4	689	4	US-09-061-769A-2
C 39	96.5	2.4	1436	2	US-08-652-971-2
C 40	96.5	2.4	1436	2	US-08-991-258A-2
C 41	96.5	2.4	1436	2	US-08-769-399-2
C 42	96.5	2.4	1436	3	US-08-991-953A-2
C 43	96	2.4	749	4	US-09-562-737-95
C 44	96	2.4	1384	4	US-08-976-255-11
C 45	96	2.4	4302	3	US-08-658-136-5

ALIGNMENTS

RESULT 1

US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LT 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 amino acids

Sequence 3, Appl1
Sequence 19, Appl
Sequence 89, Appl
Patent No. 5217870
Sequence 7, Appl1
Sequence 7, Appl1
Sequence 2, Appl1
Sequence 1, Appl1
Sequence 55, Appl
Sequence 55, Appl
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 4, Appl1
Sequence 10, Appl
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 183, App
Sequence 183, App
Sequence 183, App
Sequence 183, App
Sequence 157, App
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 95, Appl
Sequence 11, Appl
Sequence 5, Appl1

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-23

Alignment Scores:

Pred. No.: 0.00273 Length: 1185
Score: 119.50 Matches: 125
Percent Similarity: 29.91% Conservative: 44
Best Local Similarity: 22.12% Mismatches: 206
Query Match: 2.95% Indels: 191
DB: 4 Gaps: 24

US-09-743-237-1 (1-2241) x US-09-041-886-23 (1-1185)

```
QY 1991 CCCCTCCCATATCTCTGTC-----ACAGCGCTGGCTGGAAACCCCTGA 1944
DB 254 ProProThrProLysSerValSerSerGlyAlaSerGlyAlaProProThr 273
QY 1943 TGGCCACCTGACTTGTATGTGATTTGGCAGAGTCCACCTGAAGAAGAGCCTAGCTCTCT 1884
DB 274 LysProProThrProVal-GlyGlyAsnLeuProSerAlaProProAla-- 292
QY 1893 AACTTGAGCTGCTCAATAAATCTGAATGAGTACTACTAAGCAATTTGAAAGCAA 1824
DB 292 -----
QY 1823 CACATAACAAGTAGAAATACCCCTCTTTTATGAGGCGAGGAGTCCCGAGAGG 1764
DB 293 -----AsnPheProHisValThrProAsnLeuProProProAla 306
QY 1763 COTCCCCAGGACAGGAGGCGAGTGTCCATCCCGAGGCTGGCTCTGCTGCTTGCACCT 1704
DB 306 aLeuArgProLeuAsnAlaSerAlaSerProProGlyLeuGlyAla----- 322
QY 1703 TAGGATACCTGAGTAGGCTGCCAGTACCATGAGATCACTACAGCTCACAGC----- 1649
DB 323 -----GlnProLeuProGlyHisLeuProSerProTyAlaMetG1 336
QY 1648 -----TGACCTGGCCATGATCGCGGCTCTGCTCCCTCCCTGGG-- 1610
DB 336 yGlnGlyMetGlyGlyLeuProProGlyProGlyLysGlyProThrLeuAlaProSerPr 356
QY 1609 -CACAGCTTCTCGGTGCAGATTCACAGGCTGAGGCTGCCACAGGCAATTCCTTTA 1551
DB 356 OhisSerLeuPro-ProAlaSerSerAlaProAlaProProMetArgPheProTyS 376
QY 1550 CCAGCTTCACGCTACTCAATTTTCAGCCCTGGACTTGAATCGATGTGAGNATCTG 1491
DB 376 erSer-----SerSerSerSerAlaA 384
QY 1490 CGACAGGACCTTCCAAACTCTCCAGGATCATCTGCTCAGCCAGCTTGGGGAACAGTG 1431
DB 384 laAlaSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnA 404
QY 1430 CTCCTGCTCTGCTCTCCTCACCCTGGGCGCAGGAGCAGACATGTGGCTCCACTATTC 1371
DB 404 laLeuProSerTyProHisSerPhePro-----ProProThrSerL 418
QY 1370 CCAGGAGATACAGGAGGCTCCCTATTTTCTCAGTTTGTGAGTCTCTGAGACTT 1311
DB 418 euSerValSerAsnGlnProProLysTyThrGlnProSerLeuProSerGlnAlaValT 438
QY 1310 GG-----CTGGGACAAATATGCTGCT 1287
DB 438 rpSerGlnGlyProProProProProProTyGlyArgLeuLeuAlaAsnSerAsnAlaH 458
QY 1286 CTCAAAGTCCCGAGCTCCATGTAGTGGGTGTGCTCATCAGCAATTTTCTGCTGAGT 1227
DB 458 lsProGlyProPheProSerThrGlyAlaGlnSerThrAlaHisProProValSerT 478
QY 1226 TTCITCATAGTTTGTGCAAGCAATTCATTTGCAATGAAAGAAC----- 1183
DB 1183 -----
```

```
DB 478 hrHisHisHisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnH 498
QY 1182 -----ACATGATTTTGGCCTCAT-----A 1164
DB 498 isHisGlyAsnSerGlyProProProGlyAlaPheProHisProLeuGluGlyGlys 518
QY 1163 GCACTCAC-----AGTAGTTCTTCAGGAGCAGCTGAGCGCTTACA 1125
DB 518 erSerHisHisAlaHisProTyAlaMetSerProSerLeuGlySerLeuArgProTyPr 538
QY 1124 GTTG-----CACCCCTTCTGCTGTGCGAAGATTAGCAGCTC 1090
DB 538 roProGlyProAlaHisLeuProProHisSerGlnValSerTySerGlnAlaGlyP 558
QY 1089 CCAGAGGCGCTTCCCATTTTGGTTGGAAGCTTCAGGATTTCTATCAAGACAGCGCT 1030
DB 558 roAsnGlyPro--ProVal----- 563
QY 1029 TTATGGCTTTTGAAGCGCTCGAGCTCATGGCGCAGGTTTGTGCAGCTGCAGCTGTGCAGA 970
DB 564 -----SerSerSerSer-----AsnS 569
QY 969 AGTCCCGCTGGAGAAGCAGTCACAGTACCAGACAGAGATTATTTGGGTGGCCTTGC- 911
DB 569 erSerSerSerThrSerGlnGlySerTy-----ProCysS 581
QY 910 -----AGAGCGTCCAGAGGGAAGGAGCCCATTTGACAGCTGATGGA 865
DB 581 erHisProSerProSerGlnGlyProGlnGlyAlaProTyProPhe----- 596
QY 864 GAGCTCCATTATCAACCTGTGTGATGAGATTAACTTCGTGGTACTGGCAGAGATCTTC 805
DB 597 --ProProValProThr-Val-----ThrThrSerSerAlaThrLeuSerThrValIle 613
QY 804 CCATAAAGTTGTCTATTCGCTTAGGAGCTCTGACTGAGGGAAGTCTCTGCTGCTCATCGT 745
DB 614 -----AlaThrValAlaSerSerProAlaGlyTyLys 624
QY 744 ACTGAGGAAGCAGATGGAGCGCTTTGAGCTCCCTCGGCCACAGTTGTCTATGCAGAGCA 685
DB 625 ThrAlaSerProProGly-----ProProProTyGlyLysArgAlaProSer 640
QY 684 TCTGGGCGCTTCTTCAGCTGACAAATACCATGGGCTGGAGTCTTCTTCGAGAGGC 625
DB 641 ProGlyAlaTyLysThrAla-----ThrProProGlyTyLysProGlySerProPro 658
QY 624 AGCTGGAGGCTCTCTCCCTCGGATGATGGGAATTCACAGGATTCACAGGATTCCTGAGCCA 565
DB 659 SerPheArgThrGlyThrProProGlyTyArgGlyThr-----SerProPro 674
QY 564 GAGGGCTGGAAAGCTGCGTCTTCAGGGCTGCGG-----CCTGGCACACTACAC 514
DB 675 AlaGlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuPro 694
QY 513 CTGCT 509
DB 694 roAla 695
```

RESULT 2

US-08-718-388-7
; Sequence 7, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAKI
; TITLE OF INVENTION: GENE ENCODING Igg Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA

COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,388
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0230-111
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2594 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-388-7

Alignment Scores:
Pred. No.: 0.00421 Length: 2594
Percent: 119.50 Matches: 122
Percent Similarity: 31.83% Conservative: 55
Best Local Similarity: 21.94% Mismatches: 162
Query Match: 2.95% Indels: 218
DB: Gaps: 34

US-09-743-237-1 (1-2241) x US-08-718-388-7 (1-2594)

QY 1479 TTCACAACTCTCCAGGATCATCTGCT-----CAGCCAAAGCTTGGGAAACAGT----- 1432
Db 1401 PheGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu 1420
QY 1431 -----GCTCTGCTGCTCTTCTCACCCTGGCCAGCAGC----- 1396
Db 1421 GluValProAspSerProCysLeuProProThrProCysProProGlySerGluAsp 1440
QY 1395 -----AGGCACATGTGGCTCCACTTCTCCAGGAGATACAGGAGAACGCTGCC 1345
Db 1441 CysIleProSerHisLysCysProProGluLeuGluLysLysTrpGlnLysGlu----- 1458
QY 1344 TATTTTCTCAGTTTGGAGGCTCTGAGAACTTGGCTGGGGACAATAATGCTGCTCT 1285
Db 1459 -----GluPheCysGlyLeuLeuSerSerProThrGlyProLeuSerSerCysHis 1475
QY 1284 CAA-----AGTCCCAAGCTCCA-----TCTAGTGGG--- 1258
Db 1476 LysLeuValAspProGlnGlyProLeuLysAspCysIlePheAspLeuCysLeuGlyGly 1495
QY 1257 -----GTGCTCTCATCAGCATTTTCTGCTGAGCATTTCTTCATAGTTT 1213
Db 1496 GlyAsnLeuSerIleLeuCysSerAsn-----IleHisAlaTrp 1508
QY 1212 TGAAGCAATGC-----ATTTGCAATGGAAGACACATGATTTGG 1171
Db 1509 ValSerAlaCysGlnAlaAlaGlyGlyHisValGluProTrpArgThrGluThrPheCys 1528
QY 1170 CCTCATAGCACTCACAGTAGTTTCTCAGGAGCCTGAGCCTTACAGTTGACCCCTTGC 1111
Db 1529 ProMetGluCysProProAsnSerHisTrpGluLeuCysAlaAspThrCysSerLeuGly 1548
QY 1110 TGTGTCGAAGTT-----TAGCAGCTCCAGAGCGCCTT----- 1078
Db 1549 CysSerAlaLeuSerAlaProProGlnCysGlnAspGlyCysAlaGluGlyCysGlnCys 1568
QY 1077 ---TCCCAATTTTGGTTGAAAGCTTCAGGATTTCTATCAAGACAGCGCCTTTATGCTT 1021

Db 1569 AspSerGlyPheLeuTrp----- 1574
QY 1020 TGAAGCGCTCAGCTCATGCGCAGGTTGTTGC-----AGCTCAGCTGTGTCAGA 970
Db 1575 -----AsnGlyGlnAlaCysValProIleGlnGlnCysGlyCysTrpHis 1589
QY 969 AGTCCCGCTGGAGAAGCAGTACAGTACCACAGACAGAGTTATTTTGGTGGCCCTTGCA 910
Db 1590 Asn----- 1590
QY 909 GAGCAGGTCCAGAGGAAAGGAGCCCATTTGACAGCTGATGGAGAGCTCCATTATCAA 850
Db 1591 -----GlyValTyTyTyGlu 1595
QY 849 CTTGTGTGATGATTTAACTTCGCTGGTACTGGCAGAACTCTCTGGTCTACTGAGGAAGCAGAT 790
Db 1596 Pro-----GluGlnThrValLeuIle----- 1602
QY 789 TTGGCTTAGGAGCTCTGACTGAGGAAACTGCTCTGGTCTACTGAGGAAGCAGAT 730
Db 1603 -----AspAsnCysArgGlnGlnCysThrCysHisAlaGly 1614
QY 729 GGAGCGCTTTGAGCTCCTCGCCGACAGTTGT-----CTATCAGAGCATCT 682
Db 1615 LysGlyMet---ValCysGlnGluHisSerCysLysProGlyGlnValCysGlnProSer 1633
QY 681 GGGCGCTCTCTTCA---GCTGACAAATCACCATGGGGCTGGAGTCTTTCTCCGAGGCG 625
Db 1634 GlyGlyIleLeuSerCysValThrLysAspPro-----CysHisGly 1647
QY 624 ---AGCTGGAGGCTCTCT-----CTGCCCTCTGGGATGATGGAACTTGCACACAG 577
Db 1648 ValThrCysArgProGlnGluThrCysLysGluGlnGlyGlnGlyValCys----- 1665
QY 576 ATTCTCAGCAGCAGGCGCTCGGAAAGCTCGCTTTCAGGCGCTGCCCGCTG----- 526
Db 1666 LeuProAsnTrpGluAlaThr-----CysTrpLeuTrpGlyAsp 1678
QY 525 GCACACTACACCTGCTTC-----TTTGATTC---AACACTTTTCTGCTT 481
Db 1679 ProHisTrpHis-SerPheAspGlyArgLysPheAspPheGlnGlyThrCysAsnTrpVal 1698
QY 480 GCTTGTGCTCTGGACCTGGCAGTGTATCATCGGAGTCTACCAAAATG-ATTTCCCC 422
Db 1698 IleAlaThrThrGlyCys-----ProGlyValSerThrGlnGlyLeuThrPr 1714
QY 421 CTGTACACCTAGCAGCTCAGGATGTAGAGCAGGAGATCAGAGTTCTCCTAAGTTATAT 362
Db 1714 OpeThrValThrThrLysAsnGlnAsnArgGlyAsnProAlaVal-----SerTrpVal 1732
QY 361 AGCAATTTCAAGAAATGTATATAGCTTAGCAATATG---GTTGAGAAGCAGAAATTCAT 305
Db 1732 LArgValValThr-ValAlaAlaLeuGlyThrAsnIleSerIleHisLysAspGluIleG 1752
QY 304 CAAACTCTACATAGT-----TTAAGAACTCTCCCTTATCTACAGAGGGGGAGAT 251
Db 1752 LysValArgValAsnGlyValLeuThrAlaLeuProValSerValAlaAspGlyArgI 1772
QY 250 CATCAACTGAAAAAATCCTCAATTAAGTCTTTGGTGTGAAGAAAAAATCTCCC 191
Db 1772 LeuSerVal----- 1774
QY 190 AGCCAAACCCAACTACAGGCTGAAGACTATCTGCAGAACACGCAAAAGCCCTGGGTT 131
Db 1775 -----ThrGlnGlyAlaSerLysAlaLeu----- 1782
QY 130 CTATCCCTAGCCTCTGCTGGGGCGGAGGTGGGT-----GGAGTCA 92
Db 1783 --LeuValAlaAspPheGlyLeuGlnValSerTrpAspTrpAsnTrpArgValAspValT 1802
QY 91 CAAGTCCAACAGGCGCAGGAACACACACTGTGTATCTCCCTCTGCTT 48

Db 1802 hrLeuProSerSerTyHisGlyAlaValCysGlyLeuCysGly 1816
RESULT 3
US-08-718-388-9
; Sequence 9, Application US/08/18388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IGG Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718.388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-388-9

Alignment Scores:
Pred. No.: 0.00633 Length: 5405
Score: 119.50 Matches: 122
Percent Similarity: 31.8% Conservative: 55
Best Local Similarity: 21.94% Mismatches: 162
Query Match: 2.95% Indels: 218
DB: 4 Gaps: 34

US-09-743-237-1 (1-2241) x US-08-718-388-9 (1-5405)
QY 1479 TTCGAACCTCTCCAGGATCATCTGCT-----CAGCCAAAGCTTGGGAACAGT----- 1432
Db 1401 PheGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu 1420
QY 1431 -----GCTCCTCTCTGCTTCTCCTCCTCCTGCGCCAGCAGC----- 1396
Db 1421 GluValProAspSerProCysLeuProThrProCysProProGlySerGluAsp 1440
QY 1395 -----AGGCACATGTGGCTCCACTTCCAGGAGATACAGGAGAGGCTGCC 1345
Db 1441 CysIleProSerHisLysCysProGluLeuGluLysTyrglnLysGlu----- 1458
QY 1344 TATTTTCTCAGTTTGGAGTCTGAGAACTTGGCTGGGGACAAATAATGGCTGCTCT 1285
Db 1459 -----GluPheCysGlyLeuLeuSerSerProThrGlyProLeuSerSerCysHis 1475
QY 1284 CAA-----AGTCCCGCCAGCTCCA-----TGTAAGTGG----- 1258
Db 1476 LysLeuValAspProGlnGlyProLeuLysAspCysIlePheAspLeuCysLeuGlyGly 1495

QY 1257 -----GTGTGCTCATCAGCATTTTTCGTCTGAGCTTCTTCATATGTTT 1213
Db 1496 GlyAsnLeuSerIleLeuCysSerAsn-----IleHisAlaTy 1508
QY 1212 TGAAGCAATGC-----ATTTCAAATGGAAGAACACATGATTTTGG 1171
Db 1509 ValSerAlaCysGlnAlaAlaGlyGlyHisValGluProTrpArgThrGluThrPheCys 1528
QY 1170 CTTATACACTCAGCTAGTTCTTTCAGCAGCCCTGAGCGCTTACAGTTGACCCCTTGGC 1111
Db 1529 ProMetGluCysProProAsnSerHisTyrgluLeuCysAlaAspThrCysSerLeuGly 1548
QY 1110 TGTGTGCAAGTT-----TAGCAGCTCCAGACGGCCTT----- 1078
Db 1549 CysSerAlaLeuSerAlaProProGlnCysGlnAspGlyCysAlaGluGlyCysGlnCys 1568
QY 1077 ---TCCCATTTTGGTGGAAAGCTTCAGGATTCTATCAAGACACGCCCTTATGGCTT 1021
Db 1569 AspSerGlyPheLeuTy----- 1574
QY 1020 TGAAGCGCTCGAGCTCATGGCGCAGGTTGTTGC-----AGCTGCAGCTGTTGCAGA 970
Db 1575 -----AsnGlyGlnAlaCysValProIleGlnGlnCysGlyCysTyHis 1589
QY 969 AGTCCCGCTGGAGAAGCAGTCACAGTACCCAGACAGAGTTATTTTGGTGGCCCTTGCA 910
Db 1590 Asn----- 1590
QY 909 GAGCAGGTCCAGAGGGAAGGAGCCCGCCATTCACAGCTGATGGGAGAGCTCATATCAA 850
Db 1591 -----GlyValTyTrpGlu 1595
QY 849 CCTGTGTGATGAGATTAACTTCGCTGCTACTGCGCAGAAAGTCTTCCCACTAAAGTTGCA 790
Db 1596 Pro-----GluGlnThrValLeuile----- 1602
QY 789 TTGGCTTAGGAGCTCTGACTGAGGAAACTGCTGTGCTCATCTGCTGAGGAGAGCAGAT 730
Db 1603 -----AspAsnCysArgGlnGlnCysThrCysHisAlaGly 1614
QY 729 GGAGCGCTTTGAGCTCCCTCCGCGCCACAGTTGT-----CTATGAGCAGCATCT 682
Db 1615 LysGlyMet---ValCysGlnGluHisSerCysLysProGlyGlnValCysGlnProSer 1633
QY 681 GGGCGCTCTCTTCA---GCTGACAAATCACCATGGGCTGGAGCTTCTTCTCCAGGAGC 625
Db 1634 GlyGlyIleLeuSerCysValThrLysAspPro-----CysHisGly 1647
QY 624 ---AGCTGGAGCGCTCCT-----CTGCTCTCTGGATGATGGGAACCTTGCAACAGG 577
Db 1648 ValThrCysArgProGlnGluThrCysLysGluGlnGlyGlyGlnGlyValCys----- 1665
QY 576 ATTCTGAGCCAGAGGGCGCTGGAAAGCTGCTTTCAGGCGTCCCGCTG----- 526
Db 1666 LeuProAsnTyrgluAlaThr-----CysTrpLeuTrpGlyAsp 1678
QY 525 GCACACTACCACTCCTTTC-----TTTGATTTC---AACACTTTTCTTCTGCTT 481
Db 1679 ProHisTyHis-SerPheAspGlyArgLysPheAspPheGlnGlyThrCysAsnTyVa 1698
QY 480 GCTTGTCTGCTGGGAACCTGGCAGTAGTTCATCGAGCTCTCACCACAAATG-ATTTCGCC 422
Db 1698 IleuAlaThrThrGlyCys-----ProGlyValSerThrGlnGlyLeuThrPr 1714
QY 421 CTGTACACCTAGCCTCAGGATGTAGAGGAGGAGGAGTTCAGAGTTCTTCCTAAGTTATAT 362
Db 1714 oPheThrValThrThrLysAsnGlnAsnArgGlyAsnProAlaVal-----SerTyVa 1732
QY 361 AGCAATTTCAAGGAATTCCTATATAGCTTAGCAATATG---GTTGAGAGAGCAGAATTCAT 305
Db 1732 lArgValValThr-ValAlaAlaLeuGlyThrAsnIleSerIleHisLysAspGluIleG 1752
QY 304 CAAAACTCTACATAGGT-----TTAAAGAACTGTCCCTTATCTTACAGAGGGGAGAT 251


```
QY 1748 -----CCTGTCCTGGG 1759
Db 2623 SerSerThrLeuLysGlyLysThrAsnGlyAlaAspProValProGly 2638
RESULT 6
US-08-660-963-12
; Sequence 12, Application US/08660963
; Patent No. 5852187
; GENERAL INFORMATION:
; APPLICANT: Thorner, Michael O.
; APPLICANT: Gaylinn, Bruce D.
; APPLICANT: Horikawa, Reiko
; APPLICANT: Lyons Jr., Charles E.
; TITLE OF INVENTION: MOLECULAR CLONING OF THE OVINE PITUITARY
; TITLE OF INVENTION: GROWTH HORMONE RELEASING HORMONE RECEPTOR
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: POPHAM, HAIK, SCHNOBRICH & KAUFMAN, LTD.
; STREET: Metropolitan Square Building, Suite 800, 1450
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,963
; FILING DATE: 12-JUN-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Shaughnessy, Brian P.
; REGISTRATION NUMBER: 32,747
; REFERENCE/DOCKET NUMBER: 18046.036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-824-8000
; TELEFAX: 202-824-8199
; TELEX: 248516
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 498 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-660-963-12
Alignment Scores:
Pred. No.: 0.0079 Length: 498
Score: 113.00 Matches: 113
Percent Similarity: 32.45% Conservative: 46
Best Local Similarity: 23.06% Mismatches: 184
Query Match: 2.79% Indels: 147
DB: 2 Gaps: 25
US-09-743-237-1 (1-2241) x US-08-660-963-12 (1-498)
QY 2084 AAGGTTGGACAGGACACAAGTTTCTCACTTCA-----AGCAGAAATACAGGAAGA 2034
Db 18 ArgIleArgHisGluLeuAlaAlaValThrGlyAspGlyClyArgSerThrGlyPro 37
QY 2033 AAAAAAGCGAAAATCCAAACAAACCCAGCTCTAACCCGACCCCTCCCATATCTCT 1974
Db 38 ProArgLeuThrGlnAlaAlaValAlaGluLeu-----ProProTrpAlaAla 53
QY 1973 GTCACAGGCTGGCCCTGGAAAAACCCCTGATGGCCACCTGACTTGATGTGATGGGCA 1914
Db 54 GlyCysGlyAlaProAlaSerAlaCysTrpAlaPro---CysGlnSerSerTrpAla 72
```

```
QY 1913 GAGTCACCTGAAGAAGAGCGCTAGCTCTCTAACTTGAGGCTGCTCATATAACTTGAAT 1854
Db 73 ThrCysThr-----GlnSerValMetSerSerLeu----- 82
QY 1853 GGAGGTACTACTAAGCAATTTGAAAGCAACACATCAACAAAGTAGAAAAATCACCCCTCT 1794
Db 83 -----SerGluArgThrSerLysHisValTyLysLeuLeuLysGly----- 96
QY 1793 TTTTATGACAGGAGGAGTCCCCAGAGGGCTCCCCAGGAGGAGGAGGCGGCTGTCCA 1734
Db 97 -----CysPro 98
QY 1733 TCCAGGCTGGCTC-----TGCTGCTTGCACCTTTAGGATACC 1695
Db 99 ThrProProTrpAlaAlaProGlySerGlyThrGlyCysCysAlaGlyArgTrpGlnAla 118
QY 1694 TTGAGTAGCTGGCCAGTACCAGATCACTACAGACCTCACAGCTGACCTGGCCATGA 1635
Db 119 LeuGluSerGlyAlaSerProAlaArgLeuSerSerLeuThrSerAlaTrpSerGlnGly 138
QY 1634 TCGCGGCGCTCTGCTCTCCCTGGGACACAGCTTCC-----TCGGTCGAGATTCCAGGGCT 1578
Db 139 LeuArgGlyThrAlaProLeuGlnAlaGlyArgSerProSerArgLeuLeuProArgPro 158
QY 1577 GAGGCTTGCCACAGGCATTCCTTTTACCAGCTTGCCAGCTACTCAATTTTCAGCCCTT 1518
Db 159 AlaLeuCys-----Pro-TrpSerCysLeuArgArgAsnProThrSerProAr 174
QY 1517 GGACTTGAACATCGATGTGGAGAACTCTGCAGACGACCTTCCAAACTCTCCAGGATCAT 1458
Db 174 gGlySerSerThrProTrpAlaThrAlaSerArgLeuGlnProSerTrp-ProSerSerS 194
QY 1457 CTGCTCAGCCCAAGCTTGGGACAGCTGCTCTGCTCTGCTCTCCTCACCCTGGCCACGAG 1398
Db 194 erTrpSerLeuSerGlyGlySerThrAlaProGlyThrThrSerThrProSerCysSer- 213
QY 1397 GCAGGCACATGTGGCTCCACTACTTCCAGGAGATAC-----AGGAGAAGCCCTGCCT 1344
Db 214 -----ProProLeuSerSerArgArgGlnLeuCysSerArgThrProProS 229
QY 1343 ATTTTCTCAGTTTGGAGGCTCCTGAGAACTTGGCTG----- 1306
Db 229 erPheThrGly-----ArgThrTrpThrAlaAlaSerProLeuSerC 244
QY 1305 -----GGGCAATAATATGCTGCTCTCAAA 1281
Db 244 ysAlaArgLeuLeuProProLeuLeuSerArgProProThrSerAlaGlyCysTrpGlnL 264
QY 1280 GT-----CCCGAGGCTCCATGTAGTGGGTGTGCTCATCAGCAT 1242
Db 264 ysLeuCysThrProAlaSerProProHisCysProAlaGlnGlySerSerGlyGlyT 284
QY 1241 TTTTCTGCTCTG-----GACTTTCTTCATAGTTTTCAGCAAGCAATGCATTTGCAATGGA 1188
Db 284 rpPheSerLeuProGlyGlyPheLeu-----CysSerSerProAlaCysGlyTrpV 301
QY 1187 AGAACACATGATTTGGCCT----- 1168
Db 301 alaAlaSer-----TrpProLeuLysMetLeuArgAlaGlyThrTrpThrAlaProP 319
QY 1167 -----CATAGCACTCACAGTAGTTCTTCA-----GGCAGCCTGAGCGCTTACA 1125
Db 319 roThrGlySerSerLysAspProSerSerSerLeuLeuGlyThrLeuGlyPhePheS 339
QY 1124 GT---TGCACCTTTGCTGTCGAAGTTTAGCAGCTCCAGACGCGCTTTCCCAATTTT 1068
Db 339 erIleLeuSerValSerCysGlyAsnTrpSerGlnLeuArgAlaAlaSerThrProSerL 359
QY 1067 TGGTTGGAAGCTTCAGGATTTCTATCAAGACACGCGCTTTATGGCTTTTGAAGGCTCGAG 1008
Db 359 eu-----SerThrGlyValSerLeu---SerGlnArgP 369
QY 1007 CTCATGGCGCAGGCTTGTTCG-----AGCTGCAGCTGT----- 976
```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

792 QY 792 -----ACAACTTTAGTGGGAAGACATTCTGCGCAGTACAGGACGAAGTTAAATCTTCATC 842
124 Db 124 LeuCysAsnThrPheTyrValThrPheLeuPheProLeuGluThr---LeuGlnIleLeu 142
843 QY 843 ACACAGGTTGATAAATGGAGCTCTCCCATCAGCTGTCAAATGGGCTGCTTTCCCTCCCTGGA 902
143 Db 143 ThrVal-----GlyMetIleSerSerGlyValAlaSerThrAlaThrPheGlyGlyGly 159
903 QY 903 CCTGCTCTGCAAGGGCCACCCAAAATAACTCTGTCTGGGTACTGTGACTGCTTCTCTCCA-- 960
160 Db 160 ArgSer--GlyGlySerGluProValAla-CysLeuGlnGlnAlaAlaSerThrProAl 178
961 QY 961 -----CGGGGACCTTCTCCACACAGCTGCAGCTGCCAACACACTG 998
178 Db 178 aSerCysIleArgProThrAsnAlaGlyValLeuSerThrThrProSerGlyLysSerVa 198
999 QY 999 CGCCATGAGCTCGAGCGCTTAAAGCCATAAAGGGGTCTTGTATAGAAATCCCTGAAGCT 1058
198 Db 198 lGlyGluAlaHisSerValSerProProArgArgGlyValThrSerValIleLysLe 218
1059 QY 1059 TTCCAACCAAAATGGGAA-----AGGCGCTCTGGGAGCTGCT 1097
218 Db 218 uLeuSerLeuLeuTriP-LysHisValaspCysAlaArgAlaArgProThrGlySerCys- 237
1098 QY 1098 AAACCTTCGACACAGCAAGGGTGCAACTGTAGCGCTCAGCTGCCTGCAAGAACAATCTGT 1157
238 Db 238 --ThrProGluGlnGlnGlyIle-----LeuGluLysGluLeuLeu- 250
1158 QY 1158 GAGTGTCTATGAGGCCAAAATCATGTGTCTTCCATTTCGAAATGCATTGCTTGCAAAAC 1217
251 Db 251 -----ValArgTyrLeuGluGlnA 257
1218 QY 1218 TATGAGAAGATCCAGACAGCAAAATGCTGATGACACACACCCCTACTACATGGAGCGCTGG 1277
257 Db 257 rgArgGlyLysSerArgAlaIleGlyCysAspGluValThrPro----- 271
1278 QY 1278 GACTTTGAGAGCAGCCATTATTTTGCCAGCAGCAAGTTCTCAGGACCTCCAAAACCTCAG- 1336
272 Db 272 -----PheCysProThr-----ThrSerGlyThrAspp 281
1337 QY 1337 -----AAAAAATAGCAGGCGCTTCTCTGTATCTCTGGGAAGTAGTGGAGGCC 1385
281 Db 281 heProSerLeuGlnSerLysAlaGlyLeuIleSerValAsnSerGlyAlaProAlaSerH 301
1386 QY 1386 ACATGTCGCTGCTGCTGCGGCCCA 1408
301 Db 301 isGluCysAlaProTriPValPro 308

RESULT 9
US-08-281-248-2
; Sequence 2, Application US/08281248
; Patent No. 5650500
; GENERAL INFORMATION:
; APPLICANT: Raz, Abraham
; APPLICANT: Nabl, Ivan R.
; APPLICANT: Otto, Thomas
; APPLICANT: Watanabe, Hideomi
; TITLE OF INVENTION: Method of Determining Metastatic
; TITLE OF INVENTION: Potential of Tumor Cells
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DYKEMA GOSSETT
; STREET: 400 Renaissance Center
; CITY: Detroit
; STATE: MI
; COUNTRY: USA
; ZIP: 48243
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

[illegible]

```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-06421-1

Alignment Scores:
Pred. No.: 0.00699 Length: 323
Score: 112.50 Matches: 86
Percent Similarity: 36.68% Conservative: 42
Best Local Similarity: 24.64% Mismatches: 122
Query Match: 2.81% Indels: 99
DB: 5 Gaps: 18

US-09-743-237-1 (1-2241) x PCT-US93-06421-1 (1-323)
QY 444 TCCGATGAAGTACTGCCAGGTTCCCAAGCAGCAGCAAGCAAGAAAGTGTGAAATC 503
Db 33 SerGluSerPheLeuPro--SerGluGlyAlaSerSerAspProValThrLeuArgArg 51
QY 504 AAAGAAGCAGGTGGTAGTGTGCCAGGCGCAGCCCTGAAGACGACGCTTTCAGGCCOCT 563
Db 52 ArgMetLeuAlaAlaAlaArgAsnGlyGlyPheArgSerArgProProSerAlaPro 71
QY 564 CTGGCTCAGGATCTGTGTGCAAGTTCCTCATCCAGGAGCGCAGAGCCCTCCAGC 623
Db 72 Leu-----ProSerSer-----AlaAlaSerCysAlaLeu 81
QY 624 TCCGCT-----CGGAAGAAAGACTCCAGCCCATGGTGTATTTGTGAGCTGAAGCA 674
Db 82 CysProThrAspTrpArgArg-----ProValProIleLeuProLeuHisGly 97
QY 675 GCGGCCAGACTCTGTCATAGACAACCTGTGCGCGAGGAGCTCAAGCGCTCCATCTG 734
Db 98 LysAla-----GlyLeuThrAla 103
QY 735 CTTCTCTAGTACGATGACCAGAGCAGTGTCCCTCAGTCAGAGCTCCTAAGCCAATG--- 791
Db 104 LeuProLeuTyfLysAlaCysGlyLeuIleValPheGlyGlnLeuIleAsnLeuIleLeu 123
QY 792 -----ACAACTTTAGTGGAAAGACTTCTGCCAGTACCAGCAAGTTAAATCTCATC 842
Db 124 LeuCysAsnThrPheTyfValThrPheLeuPheProLeuGluThr-----LeuGlnIleLeu 142
QY 843 ACACAGGTGTGAATAGGAGCTCTCCATCAGCTCAATGGCGCTCCCTTCCCTCTGA 902
Db 143 ThrVal-----GlyMetIleSerSerGlyValAspTrpThrAlaTrpGlyGly 159
QY 903 CTGCTCTGCAAGGCCACCAATAACTCTGTCTGGTACTGTGACTGTCTCTCCA-- 960
Db 160 ArgSer---GlyGlySerGluProValAla-CysLeuGlnGlnAlaAlaSerThrProAl 178
QY 961 -----GCGGGGACTTCTGCAACAGCTGCGAGCTGCAACACCTG 998
Db 178 aserCysIleArgProThrAsnAlaGlyValLeuSerThrThrProSerGlyLysSerVa 198
QY 999 CCGCATGAGCTCGAGCGCTTCAAAGCCATAAAGGGGTGCTTTGATAGAAATCCTGAAGCT 1058
Db 198 lGlyGluAlaHisSerValSerProProArgArgGlyValThrSerValIleLysLe 218
QY 1059 TTCCACACCAAAATGGGNA-----AGGCGCTCTGGGAGCTGCT 1097
Db 218 uLeuSerLeuLeuTrp-LysHisValAspCysAlaArgAlaArgProThrGlySerCys- 237
QY 1098 AAACCTTCGACACAGAAAGGGTGCAACTGTAAAGCGCTCAGGCTGCTGAAGAACTACTGT 1157
Db 238 --ThrProGlnGlnGlyIle-----LeuGluLysGluLeuLeu- 250
QY 1158 GAGTCTATGAGGCCAAATCATGTGTCTTCCATTTCGAAATGCTTCTTGCAGAAAC 1217
Db 251 -----ValargTyfLeuGluGlnA 257
QY 1218 TATGAAGAAAGTCCAGACGCAAAATGTGTGATGACACACCCCACTACATGAGCGCTGG 1277
Db 257 rgArgGlyLysSerArgAlaIleGlyCysAspGluValThrPro----- 271
```

```
QY 1278 GACTTTGAGAGCAGCCATTATTGTTCCCGCAGCAAGTTCTCAGGACCTCCAAAAGTGA- 1336
Db 272 -----PheCysProThr-----ThrSerGlyThrAsp 281
QY 1337 -----AAAAAATAGCAGCGCTTCTCTGTATCTCTGGAAGACTAGTGGAGGCC 1385
Db 281 heProSerLeuGlnSerLysAlaGlyLeuIleSerValAsnSerGlyAlaProAlaSerH 301
QY 1386 ACATGTGCTGCTGCTGCCCA 1408
Db 301 lSgluCysAlaProTrpValPro 308

RESULT 11
US-08-751-305-2
; Sequence 2, Application US/08751305
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Tenner et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751.305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-305-2

Alignment Scores:
Pred. No.: 0.0116 Length: 652
Score: 112.00 Matches: 129
Percent Similarity: 26.88% Conservative: 53
Best Local Similarity: 19.05% Mismatches: 182
Query Match: 2.79% Indels: 313
DB: 2 Gaps: 30

US-09-743-237-1 (1-2241) x US-08-751-305-2 (1-652)
QY 39 GCGTCAGGCACACAGGATACACAG-----TGTTGTTCTCTGCTGCTGGACT 89
Db 19 GlyAlaGlyThrGlyAlaAspThrGluAlaValValCysValGlyThrAlaCysfyrthr 38
QY 90 TGTGACTCCACCACCTCCGCCCGCAGCGGCT----- 122
Db 39 AlaHisSerGlyLysLeuSerAlaAlaGluAlaGlnAsnHisCysAsnGlnGlyGly 58
QY 123 -----AGGATAGAACCCAGGCGCTTTTGGCTTTCTG-----CAG 158
Db 59 AsnLeuAlaThrValLysSerLysGluAlaGlnHisValGlnArgValLeuAlaGln 78
```

```
QY 159 ATAGTC----- 164
Db 79 LeuLeuArgArgGluAlaAlaLeuThrAlaArgMetSerLysPheTrpIleGlyLeuGln 98
QY 165 -----TTCAGCCTGTAGTATTGGGTTGGCTGGGA 194
Db 99 ArgGluLysGlyLysCysLeuAspProSerLeuProLeuLysGlyPheSerTrpValGly 118
QY 195 -----GATTTTTCCTTCCACCAAGACTTCCATTATGAGGATTTTTCAGT 245
Db 119 GlyGlyGluAsp-----ThrProTyrSerAsnTrpHisLysGlu----- 131
QY 246 TGATGATCCTCCCTCTGTAAAGACAGTCTCTTAAACCTATGTAGAGTTTGA 305
Db 131 ----- 131
QY 306 TGAATTCGTCTTCAACCATATTGCTAAGCATATATAGCAATTCCTTGAATTCCTATAT 365
Db 132 -----LeuArgAsnSerCysIleSerLys-----ArgCysValSerLeuLeu 146
QY 366 AACTTAGGAGAACCTCTGATT-----CTCCTGCCTCTACATCTCCTGAGTGTAGTGTAC 419
Db 147 AspLeuSerGlnProLeuLeuProAsnArgLeuProLys----- 159
QY 420 AGGGGAAATCATTTTGGTGAGACTCCGATGAACCTACTGTCAGGTTCCTCAAGGCAGCAAG 479
Db 160 -----TrpSerGluGlyProCysGlySerProGlySerProGlySer--- 173
QY 480 CAAGCAAGAAAAGTGTGAATCAAGAACAGCAGGTGGTAGTGCCAGGC----- 530
Db 174 -----AsnIleGluGlyPheValCysLysPheSerPheLysGlyMetCysArg 189
QY 531 -----GCAGCCCTGAGAGCCAGCCTTTCAGGCCCTCTGGCTCAGCAATCC 578
Db 190 ProLeuAlaLeuGlyGlyProGlyGlnValThrTyrThrProPheGlnThrSer 209
QY 579 TGTTCGAAGTTCCTCATCCAGGAGCAGAGGAGGCTCCAGCTGC----- 626
Db 210 -----SerSerLeuGluAlaValProPheAlaSerAlaAlaAsnValAla 224
QY 627 -----CCTCGAAGAAGACTCCAGCCCTATGGTGTAGTGTAGCTGAAA----- 671
Db 225 CysGlyGluGlyAspLysAspGluThrGlnSerHisTyrPheLeuCysLysGluLysAla 244
QY 672 -----GGAGGCGCCAGATGCTCTGCATAGACAACTGTGGCGCAGG 713
Db 245 ProAspValPheAspTrpGlySerSerGlyProLeuCysValSer----- 259
QY 714 GAGCTCAAAGCGCTCCATCTCTTCTCAGTACGATGACCAGACAGTTCCTCCCTCAGTCA 773
Db 260 -----ProLysTyr----- 262
QY 774 GAGCTCCCTAAGCCATGACAACTTTAGTGGGAAGACTTCTGCCAGTACACGCAAGTTA 833
Db 262 ----- 262
QY 834 AATCTCATCACAGGTTGATAATGGAGCTCTCCCATCAGCTGTCATGGGCTGCCTTT 893
Db 263 -----GlyCysAsnPhe 266
QY 894 CCTCTGACCTGCTCTGCAAGGCCACCCAAAATACTGTCTGGGTACTGTGACTGC 953
Db 267 AsnAsnGlyGlyCysHisGln-----AspCys 275
QY 954 TTCTCCAGCGGGGACTTCTGCAACAGCTGCAACACCTGCGCCCATGAGCTCGAG 1013
Db 276 PheGluGlyGlyAspGlySerPheLeuCysGlyCys-----ArgProGlyPheArg 292
QY 1014 CGCTTCAAGCCATTAAGGCGCTGTCTGATAGAAATCTGAAGCTTTCACCAACCAAAATG 1073
Db 293 LeuLeuAspAspLeuValThrCysAlaSerArgAsnProCysSerSerProCysArg 312
```

```
QY 1074 GGAAAGGCCCTCTGGGAGCTGTAAACTTCGACACAGCAAGGGTGCAACTGTAAAGCGC 1133
Db 313 Gly-----GlyAlaThrCysValLeu 319
QY 1134 TCAGGCTGCCTGAAGAACTAC---TGTGAGTGCTATGAGCCAAAATCATGTGTCTTC 1190
Db 320 GlyProHisGlyLysAsnTyrThrCysArgCysProGlnGlyTyrGlnLeuAspSerSer 339
QY 1191 ATTTGCAAAATGCATTGCTTGCAAAACACTATCAAGAAAGTCCAGAACAAATCTCTCATG 1250
Db 340 GlnLeuAspCysValAspValAspValAspGluCysGlnAspSerProCysAlaGlnGluCysVal 359
QY 1251 AGCACACCCCACTACATACATGAGGACCTTGGGACTTTGAGAGCAGCCATTATTGTTCCCAAGCC 1310
Db 360 AsnThrPro----- 362
QY 1311 AAGTTCACAGACCTCCAAAACCTGAGAAAAAATAGGAGGCGTTCTCTCTGTATCTCTCG 1370
Db 363 -----GlyGlyPheArgCysGluCysTrp 370
QY 1371 GAAGTAGTGGAGCCACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
Db 371 ValGlyTyrGlu-ProGlyGlyProGly----- 379
QY 1431 CACTGTTCCCAAGCTTGGCTGAGCAGATGATCTCGGAGGAGTTTGAAGGTGCTGCTGCG 1490
Db 380 -----GluGlyAlaCysGlu 384
QY 1491 CAGATTCTCACATCGAGTTCAAGTCCAGGGGCTGAAAATTGAGTAGCTGAGCTGCAAGCTGG 1550
Db 384 n-----AspValAspGluCysAlaLeuGlu 392
QY 1551 TAAGGGAATGCTGTGGCAAGCCTCAGCCCTGGGAATCTGCACCCGAGGAGGAGCTGTGC 1610
Db 392 YargSerProCys-----AlaGlnGly---CysThrAsnThrAspGlySer 406
QY 1611 C-----CAGGAGGAGCAGAGGCGCGCATCATGCCAGGTCAAGCTG 1652
Db 406 rPheHisCysSerCysGluGluGlyTyrValLeuAlaGlyGluAspGlyThr--GlnCys 425
QY 1653 TGAGCTCTGAGTATCTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1688
Db 426 GlnAspValAspGluCysValGlyProGlyProGlyProLeuCysAspSerLeuCysPheAsn 445
QY 1689 ACTCAAGTATCTTAAAGTGCACAGCAGGAGCCACCCCTGG----- 1730
Db 446 ThrGlnGlySerPheHisCys---GlyCysLeuProGlyTrpValLeuAlaProAsnGly 464
QY 1731 ---GGATGGACACTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1772
Db 465 ValSerCysThrMetGlyProValSerLeuGlyProProSerGly 479
```

RESULT 12

```
5378464-2
; Patent No. 5378464
; APPLICANT: MCEVER, RODGER P.
; TITLE OF INVENTION: MODULATION OF INFLAMMATORY RESPONSES
; BY ADMINISTRATION OF GMP-140 OR ANTIBODY TO GMP-140
; NUMBER OF SEQUENCES: 32
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/320,408
; FILING DATE: 08-MAR-1989
; SEQ ID NO: 2
; LENGTH: 830
5378464-2
```

Alignment Scores:

```
Pred. No.: 0.0214 Length: 830
Score: 110.00 Matches: 128
Percent Similarity: 29.88% Conservative: 68
Best Local Similarity: 19.51% Mismatches: 232
Query Match: 2.74% Indels: 25
DB: Gaps: 35
```

US-09-743-237-1 (1-2241) x 5378464-2 (1-830)

QY 615 GCCTCCAGCTGCTCGGAGCAAGAGCTCCAGCCCATGGTATTGTCAGCTGAAGGA 674
|||||
Db 259 AlaAlaGlnCysProProLeuLys-----IleProGlu 269
QY 675 GGCCTCCAGATGCTCTGCATAGACAAGTGTGGCGGAGGAGCTCAAAGCGCTCCATCTG 734
|||||
Db 270 ArgGlyAsnMetIleCysLeuMetSerAlaLysAla----- 281
QY 735 CTTCCTCAGTACGATGACAGAGAGCTTTCCCTCAGTCAGAG-----CTC 779
|||||
Db 282 ---ProGlnTrpGlnSerSerCysSerPheSerCysGluGluGlyPheAlaLeuValGly 300
QY 780 CTAAGGCAATGACAACCTTTAGTGGAGAGACTTCTGCCAGTACCAGCGAAGTTAAATCTC 839
|||||
Db 301 ProGluValValGlnCysThrAlaSerGlyValTrpThrAlaProAlaProValCysLys 320
QY 840 ATCACAGAGTTGATAATGGAGCTCTCCCATCA-----GCT 875
|||||
Db 321 AlaValGlnCysGlnHisLeuGluAlaProSerGluGluThrMetAspCysValHisPro 340
QY 876 GTCAATGGGGTGGCTTTCCTCTGGACCTGCTCTGCAAGGCGCCACCCAAATAAATCTG 935
|||||
Db 341 LeuThrAlaPheAlaTyrGlySerSerCysLysPheGluCysGlnProGlyTyrArgVal 360
QY 936 TCTGGG-----TACTGTGAGCTGCTTCCAGCGGGAGCTTC----- 971
|||||
Db 361 ArgGlyLeuAspMetLeuArgCysIleAspSerGlyHisTrpSerAlaProLeuProIle 380
QY 972 TGCACAGCTGCAGCTGCAACAACCTGCGCCAT----- 1004
|||||
Db 381 CysGluAlaIleSerCysGluProLeuGluSerProValMetGlySerMetAspCys 400
QY 1005 -----GAGCTGAGCGCTTCAAGCCATATAAAGCGGTGCTGTATAGAAATCTCGAAGCT 1058
|||||
Db 401 SerProSerLeuArgAlaPheGlnTyrAspThrAsnCysSerPheArgCysAlaGluGly 420
QY 1059 TTCCAACCAAAATGGGAAAGCGCTGCGGAGCTGCTAAACTCGACACAGCAAGGG 1118
|||||
Db 421 PheMet-----LeuArgGlyAlaAspIle 428
QY 1119 TGCAACTGTAAGCGCTCAGGC-----TGCTGAAGAACTAC 1154
|||||
Db 429 ValArgCysAspAsnLeuGlyGlnTrpThrAlaProAlaProValCysGlnAlaLeuGln 448
QY 1155 TGTGAG----- 1160
|||||
Db 449 CysGlnAspLeuProValProAsnGluAlaArgValAsnCysSerMetProPheGlyAla 468
QY 1161 -----TGCTATAGGCGCAAAATCATGTTCTTCC 1190
|||||
Db 469 PheArgTyrGlnSerValCysSerPheThrCysAsnGluGlyLeuLeuValGlySer 488
QY 1191 ATTTGCAAAATGCAATGCTTGCAAAACATATGAAGAA---AGTCCAGACGAGAAATGCTG 1247
|||||
Db 489 ValLeuGlnCysLeuAlaThrGlyAsnTrpAsnSerValProProGluCysGlnAlaIle 508
QY 1248 ATGAGCACACCCACTACATGAGCGCTGGGAGC-----TTT 1283
|||||
Db 509 ProCysTrpPro---LeuLeuSerProGlnAsnGlyThrMetThrCysValGlnProLeu 527
QY 1284 GAGAGAGCCATTATTTGCTCCCGCCAGCAAGTTC-----TCA 1319
|||||
Db 528 GlySerSerTyrLysSerThrCysGlnPheIleCysAspGluGlyTyrSerLeuSer 547
QY 1320 GGACCTCAAAACTGAGAAAAAATAGGCGCCCTTCTCTGCTATCTCTCGGGAAGTAGTG 1379
|||||
Db 548 GlyProGluArgLeuAspCysThrArgSerGlyArgTrpThrAspSerProPrometCys 567
QY 1380 GAGGCCACA---TGTGCTGCTGCTGCCCGGAGGTGAGGAGGAGCAGCAGGAGCTGT 1436
|||||

Db 568 GluAlaThrLysCysProGluLeuPheAla-----ProGluGlnGlySerLeuSerCys 585
QY 1437 TCCCAAGCTGGCTGAGCAGATCATCTGGAGGAGTTTGGAAAGCTGCTGCGCAGATT 1496
|||||
Db 586 SerAspThrArgGlyGlu-----PheAsnValGlySerThrCysMetPheSerCys 602
QY 1497 CTCCACATCGAGTTCAAGTCCCAAGGCGGTGAAAATTGAGTAGCGTGAAGCTGTTAAAGG 1556
|||||
Db 603 ---AsnAsnGlyPheLysLeuGlu----- 609
QY 1557 GGAATGCTGTGGCAAGCCTCAGCGCTGGGAATCTGCAC-----CGAGGAGAGCTGG 1607
|||||
Db 610 ---GlyProAsnAsnValGluCysThrMetTyrCysArgHisPro 622
QY 1608 ---TGCCAGGAGGAGGAGCGCGCATCATCGCCAGGTCACTGTGAGGTCTG 1661
|||||
Db 623 SerAlaThrProProThrCysLysGlyIleAlaSerLeuProThrPro---GlyLeu 640
QY 1662 AGTGATCTGCATGTGTACTGGCCAGCTACTCAAGGTATCTTAAAGTGCACGAGCAGCAG 1721
|||||
Db 641 ---GlnCysProAlaLeuThrThrProGlyGlnGlyThrMetTyrCysArgHisPro 659
QY 1722 CCACCTCGGGGATGGACACTGGCCCTCCTGCTCCCTGGGAGGCCCTCTGGGAGCTCCCTG 1781
|||||
Db 660 GlyThrPheGly----- 663
QY 1782 CCCTGCATAAAAGAGGGTGATTTCTACTTGTGTTGTTATGTTGCTTTTCAATTGCTT 1841
|||||
Db 664 ---PheAsnThrThrCystyr----- 669
QY 1842 AGTAGTACCTCCATTCAAGTTATTATGAGCCAGCTCAAGTTAGAGAGCTAGGCTCTTCT 1901
|||||
Db 670 ---PheGlyCysAsn 673
QY 1902 TCAGGTGA-----CTCTGCCAAATCACATACAGTCAGTGCGCCATCAGGGCTT 1952
|||||
Db 674 AlaGlyGlyPheThrLeuIleGlyAspSerThrLeuSerCysArg---ProSerGlyGln 692
QY 1953 TTTCAGGC-----CAGGCCTGT-----GACAGGAGATATGGGAGG 1988
|||||
Db 693 TrpThrAlaValThrProAlaCysArgAlaValLysCysSerGluLeuGluValAsnLys 712
QY 1989 GGGTGGGGTTAGAGCTG---GGTTGTTTGGATTTTGGCTTTTCTTCTCTCTGTTAT 2045
|||||
Db 713 ProIleAlaMetAsnCysSerAsnLeuTrpGlyPhePheSerTyrGlySerIleCys--- 731
QY 2046 TCTGCTGAAGTGAAGAACTTGTCTCTGCTCCCAACCTTTTC-----TCCATAATTACT 2099
|||||
Db 732 ---SerPheHisCysLeuGluGlyGlnLeuLeuAsnGlySerAlaGlnThr 747
QY 2100 GCTGCAGGTCG-----CCTGCTGACCACTCAGTGACCT 2135
|||||
Db 748 AlaCysGlnGluAsnGlyHisTrpSerThrThrValProThrCysGlnAlaGly---Pro 766
QY 2136 CAGACACAGAGGTGAGGTGGCTTATTAT-----GCCACACTTTTGTGTTGTTT 2186
|||||
Db 767 LeuThrIleGlnGluAlaLeuThrTyrPheGlyGlyAlaValSerThrIleGlyLeuIle 786
QY 2187 GTGAGATAAACCTTTCCAGACTCCCAAAAAAATAAAAAAATAAAAAA 2234
|||||
Db 787 MetGlyGlyThrLeuLeuAlaLeuLeuArgLysArgPheArgGlnLys 802

RESULT 13

US-08-365-470-3

; Sequence 3, Application US/08365470

; Patent No. 5632991

; GENERAL INFORMATION:

; APPLICANT: Gimbrone, Jr., Michael A.

; TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses

; TITLE OF INVENTION: Thereof

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX

Db 529 eRAlaAlaArgThrCysGlyAla-ThrGlyHisTrpSerGlyLeuLeuProThr---Cys 547
Qy 1783 CTGCATAAAAGAGGGGATTTCTACTCTGTTGTTATGTTCTTTCA-----AAT 1836
Db 548 GluAlaProThrGluSerAsnIleProLeuValAlaGlyLeuSerAlaAlaGlyLeuSer 567
Qy 1837 TGCTTAGTACCTCCATTCAGTTATATGAGCCAGCCCTCAAGTTATAGAGAGCTAGGCT 1896
Db 568 LeuLeuThrLeuAlaProPheLeuLeuTrpLeuArgLysCysLeuArgLysAlaLysLys 587
Qy 1897 CTCTTCAGGTGGACTCTGCCCAATACATACAGTCCAGTGGCCATCAG 1947
Db 588 PheValProAlaSerSerCysGlnSerLeuGluSerAspGlySerTyrGln 604
RESULT 14
US-09-209-668-19
; Sequence 19, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-668-19

Alignment Scores:
Pred. No.: 0.0257 Length: 610
Score: 108.50 Matches: 122
Percent Similarity: 31.09% Conservative: 58
Best Local Similarity: 21.07% Mismatches: 199
Query Match: 2.71% Indels: 200
DB: 3 Gaps: 31

US-09-743-237-1 (1-2241) x US-09-209-668-19 (1-610)

Qy 502 TCAAGAACGACGGTAGTGTGCGAGCGCGGCCTGGAAGACGACGCTTCCAGGCC 561
Db 131 SerLysLysLysLeuAlaLeuCysTyrThrAlaAla----- 142
Qy 562 CTCTGGCTCAGGAATCCTGTGGCAAGTTCCCATCCATCCAGGAGCAGAGCCCTCCA 621
Db 143 -----CysThrAsnThrSerCysSerGlyHis--GlyGluCysValGluThrIleAsnA 160
Qy 622 GCTGCCCTCGGAAGAACACTCCAGCCCATGGTGATTTGTCAGCTGAAAGGAGCGGCC 681
Db 160 snTyrThrCysLysCysAsp-----ProGlyPheSerG 171
Qy 682 AGATGCTCTGC-----ATAGACAACCTGTGGCGGAGGAG-----C 717
Db 171 LyLeuLysCysGluGlnIleValAsnCysThrAlaLeuGluSerProGluHisGlySerL 191
Qy 718 TCAAGCCCTCCATCTGCTTCCCTCAGTACGATGACCCAGAGC-----AGTTTCC 765
Db 191 euValCysSerHisProLeuGlyAsnPheSerTyrAsnSerSerCysSerIleSerCysA 211
Qy 766 CTCAGTCAGAGCTCCCTAAGCCATGACAACTTTA-----GTGGGAAGA---C 810
Db 211 spArgGlyTyrLeuProSerSerMetGluThrMetGlnCysMetSerSerGlyGluTrps 231
Qy 811 TTCTGCCAGTACCAGCG-----AAGTTAAATCTCATCACAGGTTGATAATG 858
Db 231 eRAlaProIleProAlaCysAsnValValGluCysAspAlaValThrAsnProAlaAsnG 251
Qy 859 GAGCTCTCCCATCAGCTGCTCAATGGGCTGCCTTTCCC----- 896
Db 251 LyPheValGluCysPheGlnAsnProGlySerPheProTrpAsnThrThrCysThrPheA 271
Qy 897 -----TCGTGACCTCTCTGCAGGGCCACCCAAAATACTCTGTCTGGTACTGTG 948
Db 271 spCysGluGluGlyPheGluLeuMetGlyAlaGlnSerLeu-----G 285
Qy 949 ACTGCTTCTCCAGGGGGACTTC-----TGCAACAGCTGCAGCTGCA 990
Db 285 InCysThrSerSerGlyAsnTrpAspAsnGluLysProThrCysLysAlaValThrCysA 305
Qy 991 ACAACCTGCGCCATGAGCTCGAGCGCTTCAAAAGCCATAAAGCGCTGTCTGTAGATAATC 1050
Db 305 rGAlaValArg----- 308
Qy 1051 CTGAAGCTTTCCCAACCAAAATGGGAAAGCCGCTCTCGGA-----GCTGCTAAAC 1101
Db 309 -----GlnProGlnAsnGlySerValArgCysSerHisSerProAlaGlyGluP 325
Qy 1102 TTCACACAGCAAAAGGTGCAAC-----TGTAAGCGCTCAGGCTGCCTGAAGAACTACT 1155
Db 325 heThrPheLysSerSerCysAsnPheThrCysGluGluGlyPheMetLeuGlnGly---- 343
Qy 1156 GTGAGTGCTATGAGGCCAAATCATGTCTCTCTCC----- 1190
Db 344 -----ProAlaGlnValGluCysThrThrGlnGlyGlnTrpThrGlnGlnIleP 360
Qy 1191 --ATTTCGAATGATGCTTTCGCAAAACTATGAAGAAAGTCCAGAACGAAAAATGCTGA 1248
Db 360 roValCysGluAlaPheGlnCysThrAlaLeu---SerAsnProGluArg----- 375
Qy 1249 TGAGCACACCCCACTACATGAG-----CCTGGGAGCTTTGAGAGACGACCTATTGTT 1302
Db 376 -----GlyTyrMetAsnCysLeuProSerAlaSerGlySerPheArgTyrGlyS 392
Qy 1303 CCCAGCAAGTTCTCAGGACCTCCA-----AAACTGAGAAAAAATAGGACAG 1350
Db 392 erSerCysGluPheSerCysGluGlnGlyPheValLeuLysGlySerLysArgLeuGlnC 412
Qy 1351 CCTTCTCCTGTATCTCTCGGAAGTAGTGGAGGCCACATGCTGCTGCTGCTGGGCCAG- 1409
Db 412 ysGlyProThrGlyGluTrpAspAsnGluLysProThrCysGluAlaValArgCysAspA 432
Qy 1410 -----GTTGAGGAAG 1419
Db 432 laValHisGlnProLysGlyLeuValArgCysAlaHisSerProIleGlyGluPheF 452
Qy 1420 CAGACGAGGAGACTGTTCCCAAGTTGGCTGAGCAGATGATCTGAGGAGTTTGA- 1478
Db 452 hrTyrLysSerSerCysAlaPheSerCysGluGluGly-----PheGluLeuTyrGlyS 470
Qy 1479 -----AGGTGCTGTGCGCAGATTCACCATCGATTCAAGTCCCAAGGGGCTGA 1527
Db 470 erThrGlnLeuGluCysThrSerGln----- 478
Qy 1528 AAATTGAGTAGCTGCAAGCTGGTAAAGGGGAATCCTGTGCAAGCCTCAGCCCTGGGA 1587
Db 478 ----- 478
Qy 1588 ATCTGCACCGAGGAAGCTGG-----TGCCAGGGAGGAGCAGAGGCC 1629
Db 479 -----GlyGlnTrpThrGluGluValProSerCysGlnValValLysCysSerS 495
Qy 1630 GCGCATATGCGCCAGCTCAGCTGTGAGGCTCTCAGTGATCTGATGATCTGCTGCGCAGCTA 1689
Db 495 erLeuAlaValProGlyLysIle-----AsnMetSerCysSerGlyGluProV 511
Qy 1690 CTCAGGTATCCTAAAGTCAAGCAGGAGCCACCTGGGGATGACACTG----- 1742
Db 511 alPheGlyThrVal---CysLysPheAlaCysProGlu---GlyTrpThrLeuAsnGlyS 529
Qy 1743 -----GCCCTCTGTCTCCCTGGGAGGCCCTCTCTGGGACTCCCTGCG 1782
Db 1782 ----- 1782

Db 529 erAlaAlaArgThrCysGlyAla-ThrGlyHisTrpSerGlyLeuLeuProThr---Cys 547
QY 1783 CCTGCATAAAGAGGGTGATTTTCTACTGTTGTTGTTATGTTGTTGTTCA-----AAT 1836
Db 548 GluAlaProThrGluSerAsnIleProLeuValAlaGlyLeuSerAlaAlaGlyLeuSer 567
QY 1837 TGGTTAGTAGTACTCATTCAAGTATATATGAGCCAGCCCTCAAGTTAGAGAGCTAGGCT 1896
Db 568 LeuLeuThrLeuAlaProPheLeuLeuTrpLeuArgLysCysLeuArgLysAlaLysLys 587
QY 1897 CTCTTCAGGTGGACTCGCCCAATCACATACAGTCAGTGGCCATCAG 1947
Db 588 PheValProAlaSerCysGlnSerLeuGluSerAspGlySerTyrgln 604

RESULT 15

US-09-009-490A-89
; Sequence 89, Application US/09009490A
; Patent No. 6300491
; GENERAL INFORMATION:
; APPLICANT: Bennett and Mirabelli
; TITLE OF INVENTION: Oligonucleotide Modulation
; TITLE OF INVENTION: of Cell Adhesion
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Jane Massey Licata
; STREET: 66 East Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WORDPERFECT 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,490A
; FILING DATE: January 20, 1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 440,740
; FILING DATE: May 12, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 063,167
; FILING DATE: May 17, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 969,151
; FILING DATE: February 10, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 007,997
; FILING DATE: January 20, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 939,855
; FILING DATE: September 2, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 567,286
; FILING DATE: August 14, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISPH-0268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 810-1515
; TELEFAX: (609) 810-1454
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: no
US-09-009-490A-89

Alignment Scores:
Pred. No.: 0.0257 Length: 610
Score: 108.50 Matches: 122
Percent Similarity: 31.09% Conservative: 58
Best Local Similarity: 21.07% Mismatches: 199
Query Match: 2.71% Indels: 200
DB: 4 Gaps: 31
US-09-743-237-1 (1-2241) x US-09-009-490A-89 (1-610)
QY 502 TCAAGAAGCAGGTGTGTGTCCTCCAGGCGGAGCCCTGAAGCAGCAGCTTCAGGCC 561
Db 131 SerLysLysLeuAlaLeuAlaLeuCysTyThrAlaAla----- 142
QY 562 CTCTGGCTCAGGAATCCTGTTGCAAGTTCCCATCATCCAGGAGGAGGAGGAGCCCTCCA 621
Db 143 -----CysThrAsnThrSerCysSerGlyHis--GlyGluCysValGluThrIleAsnA 160
QY 622 GCTGCCCTCGGAAGAAGAGCTCCAGCCCTCATGTTGTTGTCAGCTGAAGAGGAGGCC 681
Db 160 snTyThrCysLysCysAsp-----ProGlyPheSerG 171
QY 682 AGATGCTCTGC-----ATAGACAACCTGTGGCCGAGGGAG-----C 717
Db 171 lyLeuLysCysGluGlnIleValAsnCysThrAlaLeuGluSerProGluHisGlySerL 191
QY 718 TCAAGGCGTCCATCTGCTCCTCAGTACGATGACCCAGAGC-----AGTTTCC 765
Db 191 euValCysSerHisProLeuGlyAsnPheSerTyrrAsnSerCysSerIleSerCysA 211
QY 766 CTCAGTCAGAGCTCCCTAAGCCCAATGACCAACTTTA-----GTGGGAAGA---C 810
Db 211 spArgGlyTyrrLeuProSerSerMetGluThrMetGlnCysMetSerSerGlyLutrp 231
QY 811 TTCTGCCAGTACCAGCG-----AAGTTAAATCTCATCACAGGTTGATAATG 858
Db 231 erAlaProfileProAlaCysAsnValValGluCysAlaValThrAsnProAlaAsnG 251
QY 859 GAGCTCTCCCATCAGCTGTCAATGGGGCTGCTTCC----- 896
Db 251 lyPheValGluCysPheGlnAsnProGlySerPheTrpAsnThrThrCysThrPheA 271
QY 897 -----TCTGGACCTGCTCTGAAGGCGCCACCCAAAATACTGTCTGTGGTACTGTG 948
Db 271 spCysGluGluGlyPheGluLeuMetGlyAlaGlnSerLeu-----G 285
QY 949 ACTGCTTCTCCAGCGGGGACTTC-----TGCAACAGCTGCAGCTGCA 990
Db 285 lnCysThrSerGlyAsnTrpAspAsnGluLysProThrCysLysAlaValThrCysA 305
QY 991 ACAACCTGCGCCATGAGCTCGAGCGCTTCAAGGCCATAAAGCGTCTCTGTATAGAAATC 1050
Db 305 rgAlaValArg----- 308
QY 1051 CTGAAGCTTTCCAACCAAAATGGGAAAGCGCGTCTGGGA-----GCTGCTAAAC 1101
Db 309 -----GlnProGlnAsnGlySerValArgCysSerHisSerProAlaGlyGluP 325
QY 1102 TTCGACACAGCAAGGGTGTCAAC-----TGTAAGCGCTCAGGCTGCCCTGAAGAACTACT 1155
Db 325 heThrPheLysSerSerCysAsnPheThrCysGluGluGlyPheMetLeuGlnGly---- 343
QY 1156 GTGAGTGTATGAGGCCAAAATCATGTGTCTTCC----- 1190
Db 344 -----ProAlaGlnValGluCysThrThrGlnGlnGlyGlnTrpThrGlnGlnIleP 360
QY 1191 --ATTGTCAAATGCAATGCTTGCATAAACTATGAAGAAGTCCAGAACGAAATCTGA 1248
Db 360 roValCysGluAlaPheGlnCysThrAlaLeu---SerAsnProGluArg----- 375
QY 1249 TGAGCACACCCCACTACATGGAG-----CCTGGGGACTTTGAGACACCCCATTTCTT 1302
Db 376 -----GlyTyrrMetAsnCysLeuProSerAlaSerGlySerPheArgTyrrGlyS 392


```
Qy 1303 CCCAGCCCAAGTCTCTCAGGACCTCCA-----AAACTGAGAAAAAATAGGCAGG 1350
Db 392 erSerCysGluPheSerCysGluGlnGlyPheValLeuLysGlySerLysArgLeuGlnC 412
Qy 1351 CCTTCTCCTGTATCTCTGGAAGTAGTGGAGGCCACATGCTGCTGCTGGCCCCAG- 1409
Db 412 ysGlyProThrGlyGluTrpAspAsnGluLysProThrCysGluAlaValArgCysAspa 432
Qy 1410 -----GGTGAGGAAG 1419
Db 432 laValHisGlnProProLysGlyLeuValArgCysAlaHisSerProLeGlyGluPheT 452
Qy 1420 CAGAGCAGGACACTGTTCCCAAGCTTGGCTGGAGCAGATGCTCTGGAGGAGTTTGA- 1478
Db 452 hrTyrLysSerSerCysAlaPheSerCysGluGluGly-----PheGluLeuTyrGlyS 470
Qy 1479 -----AGTGCTCTGGCAGATTCTCCACATCGAGTTCAAGTCCCAAGGGGCTGA 1527
Db 470 erThrGlnLeuGluCysThrSerGln----- 478
Qy 1528 AAATTGAGTAGCGTCAAGCTGGTAAGGGGAATGCCTGTGCGAAGCCTCAGCCCTGGGA 1587
Db 478 ----- 478
Qy 1588 ATCTGCACCGAGGAGCTGG-----TGCCCGAGGAGGAGCAGAGGCC 1629
Db 479 -----GlyGlnTrpThrGluGluValProSerCysGlnValLysCysSers 495
Qy 1630 GCGCATATGCGCAGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1689
Db 495 erLeuAlaValProGlyLysile-----AsnMetSerCysSerGlyGluProv 511
Qy 1690 CTCAGGTATCCTAAAGTGCAGAGCAGAGCAGACCCTGGGGATGGACACTG----- 1742
Db 511 alPheGlyThrVal---CysLysPheAlaCysProglu---GlyTrpThrLeuAsnGlyS 529
Qy 1743 -----GCCCTCTGTCCTCTGGGAGGCCCTCTGGGACTCCCTGC 1782
Db 529 erAlaAlaArgThrCysGlyAla-ThrGlyHisTrpSerGlyLeuLeuProThr---Cys 547
Qy 1783 CCTGCATAAAGAGGGTGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1836
Db 548 GluAlaProThrGluSerAsnIleProLeuValAlaGlyLeuSerAlaAlaGlyLeuSer 567
Qy 1837 TGCTTAGTAGTACCTCCATTCAAGTTATTATCAGCCAGCCTCAGTTAGAGAGCTAGGCT 1896
Db 568 LeuLeuThrLeuAlaProPheLeuLeuTrpLeuArgLysCysLeuArgLysAlaLysLys 587
Qy 1897 CTTCTTACAGTGGACTCTGCCCAATACATACAGTCAGGTGGGCATCAG 1947
Db 588 PheValProAlaSerSerCysGlnSerLeuGluSerAspGlySerTyrGln 604
```

Search completed: April 21, 2003, 11:44:30
Job time : 65.4917 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 21, 2003, 11:28:43 ; Search time 54.3587 Seconds
(without alignments)
6234.536 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 4008
Sequence: 1 tatctgtggttgcccg.....aaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 577658

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09743237/runat_21042003_111945_462/app_query.fasta_1.4494
-DB=Published Applications_AA -QMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomum62
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MTN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09743237@cgn_1_1_49@runat_21042003_111945_462
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap:
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap:
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap:
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap:
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap:
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap:
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap:
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap:
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap:
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap:
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap:
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap:
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap:
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	405.5	10.1	438	10	US-09-220-091-13
2	131.5	3.3	1192	9	US-10-189-971-18
3	131.5	3.3	1207	9	US-10-189-971-20
4	131.5	3.3	1477	9	US-10-189-971-8

5	131.5	3.3	1535	9	US-10-189-971-14	Sequence 14, Appl
6	131.5	3.3	1593	9	US-10-189-971-4	Sequence 4, Appl
7	131.5	3.3	4123	9	US-10-213-509-5	Sequence 5, Appl
8	125.5	3.1	1057	9	US-10-189-971-6	Sequence 6, Appl
9	124	3.1	1251	9	US-10-189-971-16	Sequence 16, Appl
10	124	3.1	1342	9	US-10-189-971-24	Sequence 24, Appl
11	124	3.1	1512	9	US-10-189-971-10	Sequence 10, Appl
12	124	3.1	1570	9	US-10-189-971-12	Sequence 12, Appl
13	124	3.1	1628	9	US-10-189-971-2	Sequence 2, Appl
14	119.5	3.0	5405	9	US-10-025-380-1116	Sequence 1116, Ap
15	119.5	3.0	5405	10	US-09-922-217-1116	Sequence 96, Appl
16	116	2.9	652	9	US-09-789-919-96	Sequence 2, Appl
17	113	2.8	349	9	US-09-361-630-2	Sequence 16, Appl
18	112.5	2.8	939	10	US-09-854-845-16	Sequence 6, Appl
19	112.5	2.8	1034	10	US-09-854-845-6	Sequence 8, Appl
20	112.5	2.8	1078	10	US-09-854-845-8	Sequence 12, Appl
21	112.5	2.8	1136	10	US-09-854-845-12	Sequence 5, Appl
22	111.5	2.8	4123	9	US-10-213-509-5	Sequence 20, Appl
23	109.5	2.7	732	9	US-09-977-418-20	Sequence 22, Appl
24	109	2.7	759	9	US-10-189-971-22	Sequence 36, Appl
25	108.5	2.7	610	9	US-09-802-640-36	Sequence 14, Appl
26	107.5	2.7	954	10	US-09-854-845-14	Sequence 2, Appl
27	107.5	2.7	1049	10	US-09-854-845-2	Sequence 4, Appl
28	107.5	2.7	1093	10	US-09-854-845-4	Sequence 10, Appl
29	107.5	2.7	1151	10	US-09-854-845-10	Sequence 40, Appl
30	107	2.6	838	9	US-10-020-079-40	Sequence 38, Appl
31	107	2.6	851	9	US-10-020-079-38	Sequence 36, Appl
32	107	2.6	951	9	US-10-020-079-36	Sequence 34, Appl
33	107	2.6	957	9	US-10-020-079-34	Sequence 7, Appl
34	107	2.6	1142	9	US-10-085-108-7	Sequence 2, Appl
35	107	2.6	1142	9	US-09-899-651-2	Sequence 2, Appl
36	106.5	2.6	1436	9	US-10-042-431-78	Sequence 48, App
37	106.5	2.6	1436	9	US-09-759-1308-448	Sequence 2, Appl
38	106	2.6	610	10	US-09-748-107-2	Sequence 12, Appl
39	105	2.6	547	10	US-09-779-307-12	Sequence 13, Appl
40	105	2.6	547	10	US-09-779-307-13	Sequence 30, Appl
41	103.5	2.6	863	9	US-10-020-079-30	Sequence 28, Appl
42	103.5	2.6	876	9	US-10-020-079-30	Sequence 26, Appl
43	103.5	2.6	976	9	US-10-020-079-28	Sequence 4, Appl
44	103.5	2.6	982	9	US-10-020-079-26	
45	103.5	2.6	1036	9	US-09-373-967-4	

ALIGNMENTS

RESULT 1
US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Alignment Scores: 2.76e-25 Length: 438
Pred. No.: 405.50 Matches: 132
Score: 44.37%
Percent Similarity: 57
Conservative: 57

Best Local Similarity: 30.99% Mismatches: 158
Query Match: 10.12% Indels: 79
DB: 10 Gaps: 18

US-09-743-237-1 (1-2241) x US-09-220-091-13 (1-438)

QY 419 CAGGGGAATCATTTTGTGAGACTCCGATGAATCTACTGCCAGGTTCCTCAAGGCGAGCAA 478
|||||
Db 3 GlnGlyGluIleVal-TyrGlnAspAspAspTyrTyrAspGluSerGluIleTyrAs 22
|||||
QY 479 GCAGGCAAGAAAGTTTGAATCAAGCAAGCAGGTGTGTAGTGTGCCAGGCGGAGCC 538
|||||
Db 22 pAsnTyrGluGluGluValGluPheIleGluValAsnGluGlnLeuValProHisAsnPr 42
|||||
QY 539 TGAAGACGCGAGCTTTTCAGGCGCCCTCTGGCTCAGGAATCTGTGTCAGCTTCCCATCATC 598
|||||
Db 42 O-AsnLeuGlnAlaGlnGln-----AsnArgProGlyThrSerSerMetIleG 58
|||||
QY 599 CCAGGAGGCGAGGAGGCTCCAGCTGCTCGGAAGAAAGACTCCAGCCCATGGTGAT 658
|||||
Db 58 InGlnHisAsnArgSerMetGluValAsnGlnGlyLeuValLysAspGluProIle-Asp 77
|||||
QY 659 TTGTCAGCTGAAGG-----AGGCCCGCAGATGCTCTGCATAGACAACTGGCGC 709
|||||
Db 78 ThrSerSerHisArgValTyrValProProArgProValGlnArgLysLeuTyrLys 97
|||||
QY 710 -----GAGGAGCTCAAGCGCTCCATCTGCTCTCTCAGTAC-----GA 748
|||||
Db 98 LeuPheGlnProGlyProSerThrProGlySerSer-GlnTyrThrValArgAspLeuSe 117
|||||
QY 749 TGACGACGAGCTTCCCTCAG-----TCAGAGCTCCCTAAG----- 785
|||||
Db 117 rAsnLeuSerGlySerProSerMetTyrAspArgGlnProAlaSerLeuProArgThrVa 137
|||||
QY 786 ----CCAATGACAACCTTTAGTGGAGAGCTTCTGCCAGTACCAGCGAAGTTAACTCAT 841
|||||
Db 137 IGLnProMetGlyLeuGluMetGlyAsnSerGluGlnArgLysValTyrIleAspMetly 157
|||||
QY 842 CACACAGGTGTAATGAGCTCTCCATCATGCTGTCAATGGGCTGCTTCCCTCTGG 901
|||||
Db 157 sAspHisValSerHisIleArgLeuLysThrLysLysValPheAla----- 173
|||||
QY 902 ACCTGCTCGAAGGCGCCACCAATAACT-----CTGCTGGGTACTGTGA 949
|||||
Db 174 -ProGlyGlnArgLysProCysAsnCysThrLysSerGlnCysLeuLysLeuTyrCysAs 193
|||||
QY 950 CTGCTTCTCCAGCGGAGCTTCTGCAACAGCTGCAGCTGC-----ACCAACT 997
|||||
Db 193 pCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHisAsnAlI 213
|||||
QY 998 GCGCCATGAGCTCGAGCGCTTCAAGCCATAAAGCGTGTCTGTATAGAAATCTCGAAGC 1057
|||||
Db 213 eGluTyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsnProAsnAl 233
|||||
QY 1058 TTTCCAAACAAAATGGGGAAGCGCTGTGGAGCTGCTAAACTTGA-----CACAG 1111
|||||
Db 233 aPheLysProLysIleGlyIleAlaArgGlyGlyIleThrAspIleGluArgLeuHisG 253
|||||
QY 1112 CAAAGGTCGACTGTAAGCGCTCAGGCTGCTGAGACTACTGTGAGTGTCTATGAGGC 1171
|||||
Db 253 nLysGlyCysHisLysLysSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAl 273
|||||
QY 1172 CAAATCATGTCTTCTCCATTGCAATGCAATGCTTGTGCAAAAC----- 1217
|||||
Db 273 aLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGluThrTyrAr 293
|||||
QY 1218 -----TATGAAGAAAGT-----CCAGAACGAAAAATGCTGATGAGC-- 1253
|||||
Db 293 gMetThrArgTyrLysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeuMetSerLe 313
|||||
QY 1254 -----ACACCCCACTACATGAGCGCTGGGACTTTGAGAGCAG 1291
|||||
Db 313 uThrAsnAlaSerSerThrAlaThrProAsp-----SerGlyProGlySerValValThrAs 332
|||||

QY 1292 CCATTAT-----TTGTCCCCAGCCAAAGTTCTCAGGACCTCCAAAACCTGAGAAAAA 1342
|||||
Db 332 pGluHisGlyAspAspTyrGluAspMetLeuLeuSerHisLysProLysValGluMetAs 352
|||||
QY 1343 TAGGAGGCGCTTCTCTCTGTATCTCTGG-----GAAGTAGTGGAGGCCAC 1387
|||||
Db 352 pProArgArgPhe-----ProTyrTyrMetThrThrAspGluValValGluAlaAl 369
|||||
QY 1388 ATGTGCTCCCTGCTGCGCCAGGCTGAGGAACA-----GACGAGGA 1429
|||||
Db 369 aThrMetCysMetValAlaGlnAlaGluAlaLeuAsnTyrGluLysValGlnThrGl 389
|||||
QY 1430 GCACGTGTTCCCAAGCTGGCTGAGCAGATGATCTCGAGGAGTGTGGAAGTGGCTGTC 1489
|||||
Db 389 uAspGluLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGl 409
|||||
QY 1490 GCAGATTCTC 1499
|||||
Db 409 uGlnMetIle 412

RESULT 2
US-10-189-971-18
; Sequence 18, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walker, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2002-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR FILING DATE: 2001-07-03
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 1192
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-18

Alignment Scores:
Pred. No.: 0.0227 Length: 1192
Score: 131.50 Matches: 135
Percent Similarity: 28.72% Conservative: 56
Best Local Similarity: 20.30% Mismatches: 215
Query Match: 3.28% Indels: 260
DB: 9 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-18 (1-1192)

QY 558 GCGCCTCTGGCT-----CAGGAATCCTGCTCC----- 584
|||||
Db 305 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 324
|||||
QY 585 -----AAGTTCCTCATCTCCAGGAGGCA----- 608
|||||
Db 325 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 344
|||||
QY 609 -----GAGGAGGCTCCAGCTGCCCTCGGAGAGAAGACTCCAGCCCCATGGT 656
|||||
Db 345 CysLeuCysTyrGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 364
|||||
QY 657 ATTTGTCAGCTGAAGGAGGCGCCCATGATGCTCTCATAGACAACTGTGGCGCGAGGAG 716
|||||
Db 365 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 380
|||||
QY 717 CTCAAAGCGCTCCATCTCTCTCCTCAGTACGATGACGAGCAGTTCCTCCTCAGTCAGAG 776
|||||

```
Db 381 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 396
      ::::: ||| |||||
Qy 777 CTCCTAAGCCAAATGACAACTTTAGTGGGAAGACTTCCTCCAGTACCAGCGAAGTTAAAT 836
      ||| ::::: ||| |||||
Db 397 GluProCysAsnLeuCysThrCysLeuGlyGlyPhe----- 408
      ||| ::::: ||| |||||
Qy 837 CTCATCACACAGGTTGATAATGAGCTCTCCCATCAGCTGTCAATGGGCTGCCCTTCC 896
      ||| ||| ||| ||| ||| |||
Db 409 -----ValThrCysGlyArgArgProCysGluProProGlyCysSerHisPro 424
      ||| ||| ||| ||| ||| |||
Qy 897 TCTGGACCTGCTCTGCAAGGCCACCCAAATAACTCTGCTCTGGGTACGT----- 947
      ||| ||| ||| ||| ||| |||
Db 425 LeuIlePro-----SerGlyHisCysCysProThr 434
      ||| ||| ||| ||| ||| |||
Qy 948 -----GACTGC-----TTCCTCAGCGGGAC----- 968
      ||| ||| ||| ||| ||| |||
Db 435 CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspPro 454
      ||| ||| ||| ||| ||| |||
Qy 969 -----TTCCTGCAACAGCTGCAGCTGCAACAACCTCGCGCATGAGCTCGAGCGCTTC 1019
      ||| ||| ||| ||| ||| |||
Db 455 LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln----- 470
      ||| ||| ||| ||| ||| |||
Qy 1020 AAAGCCATAAAGCGTGTCTGATAGAAATCCTGAAGCTTCCAAACCAAAATGGGAAA 1079
      ::::: ||| |||||
Db 471 -----AspGlyGluGluPheGluGlyProAlaGlySer 481
      ||| ||| ||| ||| ||| |||
Qy 1080 -----GGCGCTCTGGAGCTGCTAACTTCGA----- 1106
      ||| ||| ||| ||| ||| |||
Db 482 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 501
      ||| ||| ||| ||| ||| |||
Qy 1107 -----CACAGCAAGGGTGCAACTGTAAGCGC---TCAGGCTCG 1142
      ||| ||| ||| ||| ||| |||
Db 502 LeuProCysLysLeuGlnValThrGluArgGlySerCysProArgCysargGlyCys 521
      ||| ||| ||| ||| ||| |||
Qy 1143 CTGAG----- 1148
      ||| ||| ||| ||| ||| |||
Db 522 LeuAlaHisGlyGluHisProGluGlySerArgTrpValProProAspSerAlaCys 541
      ||| ||| ||| ||| ||| |||
Qy 1149 AACTACTGTGAGTGTATGAG-----GCCAAATCATGTGTCTTCCATT 1193
      ||| ||| ||| ||| ||| |||
Db 542 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 561
      ||| ||| ||| ||| ||| |||
Qy 1194 TCAGAA-----TGCAATGCTTGCAAAACACTATGAAGAA 1226
      ||| ||| ||| ||| ||| |||
Db 562 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 581
      ||| ||| ||| ||| ||| |||
Qy 1227 AGTCCAGAACGAAAATGCTGATGACACACCCACTACATGAGCCTGGGAC---TTT 1283
      ::::: ||| ||||| ||| |||
Db 582 GluGlyArgLys-----TyrGluProGlyGluSerPhe 592
      ||| ||| ||| ||| ||| |||
Qy 1284 GAGACAGCCATTATTGTGCCAGCCAGTTTC-----TCAGCA 1322
      ||| ||| ||| ||| ||| |||
Db 593 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 610
      ||| ||| ||| ||| ||| |||
Qy 1323 CCTCCAAAACAGAGAAAATAGGAGCGCTCTCTCTGTATCTCTGGGAAGTAGTGGAG 1382
      ||||| ||||| ||||| ||||| ||||| |||||
Db 611 ProProSerLeuArgCysHisArg-----GlnCysProSer-----LeuValGly 626
      ||| ||| ||| ||| ||| |||
Qy 1383 GCCACATGCTGCTGCTGCTGCCAGGTGAGGAGACAGACAGAGACTGTTCCCA 1442
      ||| ||| ||| ||| ||| |||
Db 627 CysProProSerGlnLeuProProGly-----ProGlnHisCysCysPro 642
      ||| ||| ||| ||| ||| |||
Qy 1443 AGCTTGGCTGACAGATGATCC----- 1464
      ||| ||| ||| ||| ||| |||
Db 643 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 662
      ||| ||| ||| ||| ||| |||
Qy 1465 -----TGGAGCAGTTTGGAAAGGTGC 1484
      ||| ||| ||| ||| ||| |||
Db 662 oProAspProCysTyrThrCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAl 682
      ||| ||| ||| ||| ||| |||
Qy 1485 CTGTGCGAGATTCTCCACATCGAGTTCA-----AGTCCAGGGGCTGAAATTTGAG 1535
      |||| ||| ||| ||| ||| |||
```

```
Db 682 acysProGluLeuSerCysProLeuSerGluArgHisThrProProGly----- 698
      ||| ||| ||| ||| ||| |||
Qy 1536 TAGCGTGCAGCTGGTAAAGGGGAATGCTGTGGCAAGCCTCAGCCTGGGAATCTGCAC 1595
      ||| ||||| ||| ||||| ||| |||
Db 699 -SerCysCysProValCysArgGluCysValValGluAlaGluGlyArgArgValAlaAs 718
      ||| ||| ||| ||| ||| |||
Qy 1596 CGAGGAAGCTGTGCCAGGAGGA-----GCAGAGCGCGCATCATGCCCA 1643
      ||| ||| ||| ||| ||| |||
Db 718 pGlyGluSerTrpArgAspProSerAsnAlaCysIleAlaCysThrCysHisArgGlyHI 738
      ||| ||| ||| ||| ||| |||
Qy 1644 GGTACAGCTGTA-----GGTCTGAGTATGTCATGAGTGGTACTGGCCAGCC 1687
      ||| ||| ||| ||| ||| |||
Db 738 sValGluCysHisLeuGluGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 758
      ||| ||| ||| ||| ||| |||
Qy 1688 TACTCAAGTATCCTAAAGTGCAGCAGCAGCAGCACCCCTGG----- 1730
      ||| ||| ||| ||| ||| |||
Db 758 lProGlnAlaAspSerCysGluArgCysGlnAlaProThrGlnSerCysValHisG1 778
      ||| ||| ||| ||| ||| |||
Qy 1731 -----GGATGGACACTG----- 1742
      ||| ||| ||| ||| ||| |||
Db 778 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 798
      ||| ||| ||| ||| ||| |||
Qy 1743 -----GCCCTCTCTCCCTGGGAGGCC 1765
      ||| ||| ||| ||| ||| |||
Db 798 smetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 816
      ||| ||| ||| ||| ||| |||
Qy 1766 CTCTGGGAGCTCCCTGCCCTGCATATAAAGAGGGTGATTTCTACTTGTGTATGCTGT 1825
      ||| ||| ||| ||| ||| |||
Db 817 ----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 830
      ||| ||| ||| ||| ||| |||
Qy 1826 TGCTTCAATATGCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCAGCCTCAAGTTAG 1885
      ||| ||| ||| ||| ||| |||
Db 831 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 848
      ||| ||| ||| ||| ||| |||
Qy 1886 AGAGCTAGGCTCT-----TCTTCAGGTGGACTCTCCCAATCATACATACTAG 1936
      ||| ||| ||| ||| ||| |||
Db 848 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 868
      ||| ||| ||| ||| ||| |||
Qy 1937 GTGGCCATCAGGGGTTTTTCCAGGCCAGCCTGTGACAGAGATATGGGAGGGGCTCG 1996
      ||| ||| ||| ||| ||| |||
Db 868 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 887
      ||| ||| ||| ||| ||| |||
Qy 1997 GTTAGAGCTGG 2007
      ||| ||| ||| ||| ||| |||
Db 887 lyValAlaTrp 890
      ||| ||| ||| ||| ||| |||
RESULT 3
US-10-189-971-20
; Sequence 20, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-20
Alignment Scores: 0.0228 Length: 1207
Pred. No.:
```

Score:	131.50	Matches:	135
Percent Similarity:	28.72%	Conservative:	56
Best Local Similarity:	20.30%	Mismatches:	215
Query Match:	3.28%	Indels:	260
DB:	9	Gaps:	36

US-09-743-237-1 (1-2241) x US-10-189-971-20 (1-1207)

QY	558	GCCTCTGGCT	-----CAGGAATCCTGTTGC-----	584
DB	320	AlaProCysAlaHisProArgGlnGlyProCysProSerCysAspGlyCysLeu	tyr 339	
QY	585	-----AAGTTCCCATCATCCACGAGGCA-----	608	
DB	340	GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu	359	
QY	609	-----GAGGAGGCCTCCAGCTGCCCTCCGGAAGAAAGACTCCAGGCCCATGTGTG	656	
DB	360	CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeu	Cys 379	
QY	657	ATTGTGTCAGCTGAAGAGGCGCCAGATGCTCTGCATAGACAACGTGTGGCGCGAGGAG	716	
DB	380	ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu	395	
QY	717	CTCAAGCGCTCCATCTGCTTCCTCAGTACGATGACCAGAGCAGTGTTCCTCCAGTCAGAG	776	
DB	396	-----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg	411	
QY	777	CTCCCTAAGCCAATAGCAAACTTTAGTGGGAGCACTTCTGCCAGTACCACGCAAGTTAAAT	836	
DB	412	GluProCysAsnLeuCysThrCysLeuGlyGlyPhe-----	423	
QY	837	CTCATCACACAGGTTGATTAATGGAGCTTCGCCATCAGCTGCATGGGCTGCCTTTCC	896	
DB	424	-----ValThrCysGlyArgArgProCysGluProProGlyCysSerHis	439	
QY	897	TCTGGACCTGCTCTGCAAGGGCCACCAAAATAACTGTCTGTGGTACTGT-----	947	
DB	440	LeuIlePro-----SerGlyHisCysCysProThr	449	
QY	948	-----GACTGC-----TTCTCCAGCGGGAC-----	968	
DB	450	CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAsp	469	
QY	969	-----TTCTGCAACAGCTGCAGCTGCCACACACCTGCGCCATGAGCTCGAGCGCTTC	1019	
DB	470	LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln-----	485	
QY	1020	AAAGCCATAAGGCGTGTCTGTGATAAAATCTGAAGCTTTCCACCAAAAAATGGGAAA	1079	
DB	486	-----AspGlyGluGluPheGluGlyProAlaGlySer	496	
QY	1080	-----GGCCCTCTGGGAGCTGCTAAACTTCGA-----	1106	
DB	497	CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysPro	516	
QY	1107	-----CACACCAAGGGTGCACCTGTAACGC-----TCAGGCTGC	1142	
DB	517	LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys	536	
QY	1143	CTGAAG-----	1148	
DB	537	LeuAlaHisGlyGluHisProGluGlySerArgTrpValProProAspSerAlaCys	556	
QY	1149	AACTACTGTGAGTGCATGAG-----GCCAAATCATGTGTCTTCCATT	1193	
DB	557	SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSer	576	
QY	1194	TGCAAA-----TGCATTGCTTGCAAAACATATGAACAA	1226	
DB	577	CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGlyHis	596	
QY	1227	AGTCCAGAACGAAAAATGCTGATGACACACCCCACTACATGAGCGCTGGGAC-----	1283	

RESULT 4

US-10-189-971-8

Qy	558	GCCCCCTCTGGCT-----CAGGAATCCTGTTGC-----	584
Db	590	AlaProCysAlaHisProArgGlnGlyProCysProSerCysAspGlyCysLeuTyr	609
Qy	585	-----AAGTCCCATCATCCAGGAGGCA-----	608
Db	610	GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu	629
Qy	609	-----CAGGAGCCCTCCAGCTGCCCTCGGAAGAAGACTCCAGGCCCATGGTG	656
Db	630	CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys	649
Qy	657	ATTGTGACGTGAAGAGGCCGCCAGATGCTCTCATAGACAACACTGTGGCGCAGGGAG	716
Db	650	ProPheProAlaArgLysAsp-----CysCysProAspCysAspGlyCysGlu	665
Qy	717	CTCAAGCGGTCNCTGCTTCTCAGTACGATGACACAGAGAGTGTTCCCTCAGTCTAGAG	776
Db	666	-----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg	681
Qy	777	CTCCCTAAGCCATGACAACACTTTAGTGGGAAGACTTCTGCCAGTACACGCAAGTTAAAT	836
Db	682	GluProCysAsnLeuCysThrCysLeuGlyGlyPhe-----	693
Qy	837	CTCATCACACAGGTTGATAATGGAGCTCTCCCATCAGCTGCAATGGGCGTGCTTTCCC	896
Db	694	-----ValThrCysGlyArgArgProCysGluProProGlyCysSerHisPro	709
Qy	897	TCTGGACCTGCTCTGCAAGGGCCACCCAAAATAACTGTCTGGGTACTGT-----	947
Db	710	LeuIlePro-----SerGlyHisCysCysProThr	719
Qy	948	-----GACTGC-----TTCTCCAGGGGGAG-----	968
Db	720	CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspPro	739
Qy	969	-----TTCTGCAACAGCTGCAGCTGCACACACCTGCGCCCATGAGCTCGACGGCTTC	1019
Db	740	LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln-----	755

```
Db 1083 smetalaglythrvalargCysGlnSerClnargCysSerProLeuSerCysGly 1101
Qy 1766 CTCTGGGAGCTCCCTGCCTGCATAAAGAGGGGATTTCTTACTCTGTGTATGTATGTT 1825
Db 1102 ----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 1115
Qy 1826 TGCTTTCAATTCCTAGTAGTACCTCCATTCAAGTATTATGAGCCAGCCCTCAAGTTAG 1885
Db 1116 ----ArgCysLeuProArgProAlaSerCysMetalaPheGlyAspProHisTyrAr 1133
Qy 1886 AGAGCTAGCTCT-----TCTTCAGGTGGACTCTGCCAATATCATACACAGTACAG 1936
Db 1133 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1153
Qy 1937 GTGGCCATCAGGGGTTTTCCAGCGCAGCGCTGTGCACAGGAGATATGGAGGGGGTCGG 1996
Db 1153 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspArgGlyArgSerG 1172
Qy 1997 GTTAGAGCTGG 2007
Db 1172 LyValAlaTrrp 1175

RESULT 5
US-10-189-971-14
; Sequence 14, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and Polynucleot
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1535
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-14

Alignment Scores:
Pred. No.: 0.0246 Length: 1535
Score: 131.50 Matches: 135
Percent Similarity: 28.72% Conservative: 56
Best Local Similarity: 20.30% Mismatches: 215
Query Match: 3.28% Indels: 260
Db: 9 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-14 (1-1535)
Qy 558 GCCCTCTGGCT-----CAGGAATCTGTTC----- 584
Db 648 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 667
Qy 585 -----AAGTTCCATCATCCAGGAGCA----- 608
Db 668 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 687
Qy 609 -----GAGGAGGCTCAGCTGCCCTCGGAGAAGACTCCAGGCCCATGTGG 656
Db 688 CysLeuCysTrrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 707
Qy 657 ATTTGTACAGTGAAGGAGCGGCCAGATGCTCTCATAGACAACTGTGGCGCGGAGGAG 716
Db 708 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 723
```

```
Qy 717 CTAAAGCGCTCCATCTCTCTCCTCAGTACGATCACCAGACAGTTCCTCCTCAGTACAG 776
Db 724 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 739
Qy 777 CTCCTTAGCCAAATGACAACTTTAGTGGGAAGACTTTCGCCAGTACGACGCAAGTTAAAT 836
Db 740 GluProCysAsnLeuCysThrCysLeuGlyGlyPhe----- 751
Qy 837 CTCATCACACAGGTGATTAATGGAGCTCTCCCATCAGCTGTCAATGGGGCTGCCTTTCC 896
Db 752 -----ValThrCysGlyArgArgProCysGluProProGlyCysSerHisPro 767
Qy 897 TCTGGACCTGCTCTGCAAGGCCACCCCAATAAATCTCTGTGGTACTGT----- 947
Db 768 LeuIlePro-----SerGlyHisCysCysProThr 777
Qy 948 -----GACTGC-----TTCCTCCAGCGGGGAC----- 968
Db 778 CysGlnGlyCysArgTyrHisGlyValThrAlaSerGlyGluThrLeuProAspPro 797
Qy 969 -----TTCGCAACAGCTGCAGCTGCAACAACCTGCCCATGAGCTCGAGCGCTTC 1019
Db 798 LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln----- 813
Qy 1020 AAGCCATAAAGCGGTCTTGTAGATAAAATCCTGAAGCTTTCCAAACCAAAAAATGGGAAA 1079
Db 814 -----AspGlyGluGluPheGluGlyProAlaGlySer 824
Qy 1080 -----GGCGTCTGGAGCTGTCTAACTTCGA----- 1106
Db 825 CysGluTrpCysArgCysGlnAlaGlyValSerCysValArgLeuGlnCysProPro 844
Qy 1107 -----CACAGCAAGGTCGCAACTCTAAGCG-----TCAGGCTGC 1142
Db 845 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 864
Qy 1143 CTGAAG----- 1148
Db 865 LeuAlaHisGlyGluGluHisProGluGlySerArgTrrpValProProAspSerAlaCys 884
Qy 1149 AACTACTGTGAGTCTATGAG-----GCCAAATCATGTGTCTTCCATT 1193
Db 885 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 904
Qy 1194 TGCAAA-----TGCAATTGCTTCAAAAACATATGAAGA 1226
Db 905 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 924
Qy 1227 AGTCCAGAACGAAAAATGCTGATGAGCACACCCCACTACATGGAGCCTGGGAC---TTT 1283
Db 925 GluGlyArgLys-----TyrGluProGlyGluSerPhe 935
Qy 1284 GAGAGCAGCCATTATTGTCTCCAGCCCAAGTTC-----TCAGGA 1322
Db 936 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 953
Qy 1323 CTCCAAAACTGAGAAAAAATAGCAGGCGCTCTCTCTATCTCTCTGGGAAGTAGTGAG 1382
Db 954 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 969
Qy 1383 GCCACATGTGCTCTGCTGGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1442
Db 970 CysProProSerGlnLeuLeuProGly-----ProGlnHisCysCysPro 985
Qy 1443 AGCTTGGCTGAGCAGATCATCC----- 1464
Db 986 ThrCysAlaGluAlaLeu-SerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaPr 1005
Qy 1465 -----TGGAGGAGTTTGGAGGTGC 1484
Db 1005 opProAspProcysTrrpThrCysGlnCysGlnAspLeuThrTrrpLeuCysIleHisGlnAl 1025
```


QY 1485 CTGTCCAGATTCTCCATCGAGTTCA-----AGTCCAAGGGGCTGAAATTTGAG 1535
Db 1025 aCysProGluLeuSerCysProLeuSerGluArgHisThrProGly----- 1041
QY 1536 TACCGTGCAGCTGTTAAAGGGAATGCTGTGGCAAGCTCAGCCCTGGGAATCTGCAC 1595
Db 1042 -SerCysCysProValCysArgGluCysValGluAlaGluGlyArgArgValAlaAs 1061
QY 1596 CGAGGAAGCTGGTCCACGGAGGA-----GCAGAGGCGCGCATCATGGCCA 1643
Db 1061 pGlyGluSerTrpArgAspProSerAsnAlaCysHisLeuArgGlyHis 1081
QY 1644 GGTACGCTCTGA-----GGTCTGAGTGTCTGCATGTGAGTGGCCAGCC 1687
Db 1081 sValGluCysHisLeuGluCysGlnAlaLeuSerCysProHisGlyTrpAlaLysVa 1101
QY 1688 TACTCAAGTATTCCTAAAGTGCAGGAGGAGCCAGCCACCTGG----- 1730
Db 1101 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisG 1121
QY 1731 -----GGATGGACACTG----- 1742
Db 1121 nGlyArgGluValAlaSerGlyGluArgTrpThrValAspThrCysThrSerCysSerCy 1141
QY 1743 -----GCCCTCCTCTCCCTGCGGGAGGCC 1765
Db 1141 sMetalAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 1159
QY 1766 CTCTGGGACTCCCTCCCTGCGATAAAAGAGGGTATTTCTACTGTGTGTATGTGTT 1825
Db 1160 ----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 1173
QY 1826 TGCTTTCAAATTGCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCAGCCCTCAAGTTAG 1885
Db 1174 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 1191
QY 1886 AGAGCTAGGCTCT-----TCCTTTCAGGTGACTCTGCCCCAAATCACAACAAGTCAG 1936
Db 1191 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 1211
QY 1937 GTGGCCATCAGGGTTTTCCAGGCCAGCCCTGTGACAGAGATATGGGAGGGGGTCTGG 1996
Db 1211 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 1230
QY 1997 GTTAGAGCTGG 2007
Db 1230 lyValAlaTrp 1233

RESULT 6

US-10-189-971-4
; Sequence 4, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human kielin-like Proteins and Polynucleoti
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1593
; TYPE: PRP
; ORGANISM: homo sapiens
US-10-189-971-4

Alignment Scores:
Pred. No.: 0.0249 Length: 1593
Score: 131.50 Matches: 135
Percent Similarity: 28.72% Conservative: 56
Best Local Similarity: 20.30% Mismatches: 215
Query Match: 3.28% Indels: 260
DB: 9 Gaps: 36

US-09-743-237-1 (1-2241) x US-10-189-971-4 (1-1593)

QY 558 GCCCTCTCTGGCT-----CAGGAATCCTGTTC----- 584
Db 706 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 725
QY 585 -----AAGTCCCATCATCCAGGAGCA----- 608
Db 726 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 745
QY 609 -----CAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGGTG 656
Db 746 CysLeuCysTrpGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 765
QY 657 ATTTGTCAGCTGAAGAGGCGGCCAGATGCTCTGCATACACAACTGTGGCGGAGGGAG 716
Db 766 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 781
QY 717 CTCAAAGCGCTCCATCTCTCTCAGTAGGATGACGAGCAGGAGTTTCCTCAGTCAGAG 776
Db 782 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 797
QY 777 CTCCTTAAGCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGCGAAGTTAAAT 836
Db 798 GluProCysAsnLeuLeuSerCysLeuGlyGlyPhe----- 809
QY 837 CTCATCACACAGGTTGATATGAGAGCTCTCCCATCAGCTGCTCAATGGGGTGCCTTTCCC 896
Db 810 -----ValThrCysGlyArgArgProCysGluProProGlyCysSerHisPro 825
QY 897 TCTGGACCTGCTCTGCAAGGGCCACCAATAAATACTGCTGCTGGTACTGT----- 947
Db 826 LeuIlePro-----SerGlyHisCysCysProThr 835
QY 948 -----GACTGC-----TCTCTCAGCGGGGAC----- 968
Db 836 CysGlnGlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspPro 855
QY 969 -----TTCTGCAACAGCTGCAGCTGCACAACTCGCGCCATCAGCTCGAGCGCTTC 1019
Db 856 LeuAspProThrCysSerLeuCysThrCysGlnGlyArgGluHisGln----- 871
QY 1020 AAAGCCATAAAGCGTCTTGTATGATAAATCCTGAAGCTTTCACCAACCAAAATGGGAAA 1079
Db 872 -----AspGlyGluGluPheGluGlyProAlaGlySer 882
QY 1080 -----GGCCCTCTGGGAGCTGCTAACTTCGA----- 1106
Db 883 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 902
QY 1107 -----CACAGCAAGGGTGCACACTGTAAAGCGC-----TCAGGCTGC 1142
Db 903 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 922
QY 1143 CTGAAG----- 1148
Db 923 LeuAlaHisGlyGluGluHisProGluGlySerArgTrpValProProAspSerAlaCys 942
QY 1149 AACTACTGTGAGTGCATGAG-----GCCAAATCATGTCTTCTTCATT 1193
Db 943 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 962
QY 1194 TGCAAA-----TGCATTCTTGCACAAATATGAAGAA 1226
Db 963 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 982

Db 3186 ValThrAsnCysThrAlaIleGluGlyAlaGluTyrSerProCysGlyProCysPro 3205
QY 984 ---AGCTGCAACAACTGCGCCATCAGCTCGAGCGCTTCAAAGCCATAAAGCGGTCTT 1040
Db 3206 ArgSerCysAspAspLeuValHis-----CysVal 3215
QY 1041 GATAGA---AATCCTGAAGCTTTCACCAAAA----- 1070
Db 3216 TrpArgCysGlnProGlyCysTyrCysProProGlyGlnValLeuSerSerAsnGlyAla 3235
QY 1071 -----ATGGGGAAGCGCTGGAGCTGCTAACTT-----CGACACAGC 1112
Db 3236 IleCysValGlnProGlyHisCysSerCysLeuAspLeuLeuThrGlyGlnArgHis 3255
QY 1113 AAAGGTGCAACTGAAGCGC---TCAGGCTGCCTGAAGAAGTACTGTGAGTCTATGAG 1169
Db 3256 ProGlyAlaArgLeuAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGlu 3273
QY 1170 GCCAAATCATGTGTCTTCCATTTCGAATGCTTGCATAAAGTATGAAGAAAGT 1229
Db 3274 GlyArgLeuAsnCysThrAspLeu---ProCysProAspCys----- 3286
QY 1230 CCAGAACCAAAATCTGATGAGCACACCCACTACATGGAGCCTGGGAGCTTTCAGAGC 1289
Db 3287 -----GlyGlyGlyGlnSer 3291
QY 1290 AGCCATTATTTGCTCCCGAGCAAGTTCTCAGGACCTCCAAAGTCTGAGAAAAATAGGCAG 1349
Db 3292 LeuHis-ProCysGlyGlnProCysProArgSerCysGlnAsp----- 3305
QY 1350 GCCTTCTCTCTATCTCCTGGAAGTAGTGGAGGCACATGCTGCTGCTGCTGCC- 1407
Db 3306 -----LeuSerProGlySerValCysGlnProGlySerValGlyCysGlnProTh 3322
QY 1407 ----- 1407
Db 3322 rCysGlyCysProLeuGlyGlnLeuSerGlnAspGlyLeuCysValProAlaHisCys 3342
QY 1408 -AGGCTGAGGAAGCAGACAGGAGCAGCTGTTCCCA----- 1443
Db 3342 sArgCysGlnTyrGlnProGlyAlaMetAlaProSerPheValProSerThrCysValAl 3362
QY 1444 -----GCTTGGCTGAGCAG 1457
Db 3362 aglyIleLeuGlnCysGlnGluValProAspCysProAspProGlyValTrpSerSerTr 3382
QY 1458 ATGATCCTGGAGGATTTGGAAAGTGCCTGCGCAGATTCACATCGAGTTCAAGTCC 1517
Db 3382 pGlyProTrpGluAspCysSerValSerCys-----G1 3393
QY 1518 AAGGGCTGAAATAGTAGTACGCTGCAAGCTGGTAAAGGGGAATGCCTGTGCAAGCCTC 1577
Db 3393 yGlyGlyGlnLeuArgSerArgCysAlaArgProProCysProGlyProAlaArg 3413
QY 1578 AGCCCTGGGAATCTGCACCGAGGAGCTGGTCCCGAGGAGCAGAGCGCGCATCA 1637
Db 3413 gGlnSerArgThrCysSerThrGlnValCysArgGluAlaGlyCysProAlaGlyArg-- 3432
QY 1638 TGGCCAGGTCAGCTGAGGCTGAGTGTGATGTCATGTCATGTCAGGCTACTCAAGGT 1697
Db 3432 ----- 3432
QY 1698 ATCCTAAAGTCAACGAGCAGACGACCCCTGGGAGTGGAGTGGAGTGGCTGCTGCTG 1757
Db 3433 -----LeuTyrArgGluCysGlnProGlyGluGly-----CysProPh 3445
QY 1758 G-----GGAGGCGCTCTCGGGAC 1775
Db 3445 eSerCysAlaHisValThrGlnGlnValGlyCysPheSerGluGlyCysGluGluGlyCy 3465
QY 1776 TCCCTGCCCTGCATAAAAGAGGGTATTTCTACTGTTGTATGTTGCTTTCAA 1835
Db 3465 sHisCysPro-----GluGlyThrPheGlnHisArgLeuAlaCys---ValGlnG1 3481

QY 1836 TTGCTTAGTAGTACCTCCATTCAGTTTATGACCGAGCTCAAGTTTAGAGAGCTAGGC 1895
Db 3481 uCysProCysVal-----LeuThrAlaTrpLeu--LeuGlnGluLeuGly 3495
QY 1896 TCTTCTTCAGGTGGACTCTGCCCAATCACATACAAGTCAGGTGCCCATCAGGGGTTTTT 1955
Db 3496 AlaThrIleGlyAsp-----ProGlyGln-----ProLeuGly----- 3506
QY 1956 CCAGCCAGGCTGTGACAGGAGATATGGAGGGGGTGGGTTAGAGCTGGGTTGTTT 2015
Db 3507 ProGlyAspGluLeuAspSer-----GlyGlnThrLeuArgThrSerCysGly----- 3522
QY 2016 GGATTTTTTGGCTTTTTTCTTCCTGATTTCTGCTGAAGTGAGAAAACCTTGTCTCTCG 2075
Db 3523 -----AsnCysSerCysAlaHisGlyLeuSerCysSerLeu 3535
QY 2076 TCCAACTTTTCTCCATAATTACTGCTGACGGTGCCTGTGACGACAGTCACAGTGCCT 2135
Db 3536 AspAspCysPheGlu-----AlaAspGlyGlyPheGlyPro 3547
QY 2136 CAGACACAGAGGTGAGTGGCTTATTATGCCCCACACTTTG----- 2177
Db 3548 TrpSerProTrpGlyProCysSerArgSerCysGlyGlyLeuGlyThrArgThrArgSer 3567
QY 2178 -----TGTTTTTGTGTGAGATAAACCTTCCACACTCCCA 2213
Db 3568 ArgGlnCysValLeuThrMetProThrLeuSerGluLeuPro 3581

RESULT 8
US-10-189-971-6
; Sequence 6, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucle
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10189, 971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302, 949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315, 634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-6

Alignment Scores:
Pred. No.: 0.0689 Length: 1057
Score: 125.50 Matches: 149
Percent Similarity: 27.56% Conservative: 66
Best Local Similarity: 19.10% Mismatches: 267
Query Match: 3.13% Indels: 299
DB: 9 Gaps: 37

US-09-743-237-1 (1-2241) x US-10-189-971-6 (1-1057)

QY 374 AGAACCCTCTGATTCCTGCTCTTACATCCTGAGTGTACAGGGGAAATCATTT 433
Db 41 ArgThrSerProThrProLeuThrProAlaValCysValAlaValSerGlyAsnValGln 60
QY 434 TTGGTGACACTCCGATGACTACTGCCAGGTTCCTCCAGCAGCAGCAAGCAAGAAAG 493
Db 61 CysLeuAlaArgArgCysValProLeuProCysProGluProValLeuLeuProGlyGlu 80
QY 494 TGTTCAAATCAAAGAAGCAGGTGGTAGTGTGCCAGCGCGCAGCCCTGAAGACGACGCTTT 553

Db	81	CysCysProGlnCys-ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAla	100
Qy	554	CCAG-----	557
Db	100	aHisAlaArgHisGlnGluTyrPheSerProProGlyAspProCysArgArgCysLeuCy	120
Qy	558	-----	569
Db	120	sLeuAspGlySerValSerCysGlnArgLeuProCysProProAlaPaProCysAlaHisPr	140
Qy	570	---CAGGAATCCTGTGC-----	584
Db	140	oArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyrGlnGlyLysGluPheAl	160
Qy	585	-----AAGTTCCCATCATCCAGAGGCA-----	613
Db	160	aSerGlyGluArgPheProSerProThrAlaAlaCysHisLeuCysLeuCyTyrGluGl	180
Qy	614	GGCCTCCAGCTGCCCTGGGAAGAAGACTCCAGCCCCCATGTTGATTTGTCAGCTGAAGG	673
Db	180	ySerValSerCysGlnProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGl	200
Qy	674	AGGCGCCACAGTCTGTGCATACAGAACTGTGTGGCGGAGGAGCTCAAAAGCGCTCCATCT	733
Db	200	yAsp-----CysProAspCysAspGlyCysGlu-----	212
Qy	734	GCTTCCCTCAGTACGATGACCAGAGCAGTTTCCCTCAGTCAGAGCTCCCTAAGCCCAATGAC	793
Db	212	uGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArgGluProCysAsnLeuCy	232
Qy	794	AACTTTAGTGGGAAGACTTCTGTCCA-----	818
Db	232	sThrCysLeuGlyGlyPheValThrCysGlyArgArgProCysGluProProGlyCysSe	252
Qy	819	-----GTACCAGC-----	838
Db	252	rHisProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGl	272
Qy	839	CATCACACAGTTGATAATGGAGCTCTCCCATCAGCTGTCATGGGCGTCTTT-----	893
Db	272	yValThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeuCy	292
Qy	894	-----CCCTCTGGACCTGCTCTGCAGG	916
Db	292	sThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLeuCysPr	312
Qy	917	GCCACCAAAATAACTCTGCTGGGTACTGTGACTGCTTCTCCAGCGGGAGCTTCTGCAA	976
Db	312	oHisProSer-----ProGlyProCysPheCys-----ProValCysHi	325
Qy	977	CAGCTGCAGCTGCAACAACCTGGCGCATCAGCTCGAGCGCTTCAAAAGCCATAAAGCGCTG	1036
Db	325	sSerCysLeuSerGlnGlyArgGluHisGln-----	335
Qy	1037	TCTTGATAGAAATCCTGAAGCTTTCACACCAAAATGGGAAA-----	1079
Db	336	-----AspGlyGluGluPheGluGlyProAlaGlySerCysGluTyrCysArgCy	352
Qy	1080	-----GGCGGTCTGGGAGCTGCTAACTTCA-----	1106
Db	352	sGlnAlaGlyGlnValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGl	372
Qy	1107	---CACAGCAAGGGTGCNACTGTAAAGCGC---TCAGGCTGCCTGAAG-----	1148
Db	372	nValThrGluArgGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGl	392
Qy	1149	-----AACTACTGTGAGTGTCTA-----	1165
Db	392	uHisProGluGlySerArgTyrValProProAspSerAlaCysSerCysValCysHi	412
Qy	1166	TGAG-----GCCAAATCATGTCTTCTCCATTTGCAAA-----	1199


```
Db 817 lProGlnAlaAspSerCysCysGluArgCysGlnAlaProThrGlnSerCysValHisG1 837
QY 1731 -----GGATGGACACTG----- 1742
Db 837 nGlyArgGluValAlaSerGlyGluArgTTPThrValAspThrCysThrSerCysSerCy 857
QY 1743 -----GCCTCTGTCCTGGGGAGGCC 1765
Db 857 sMetAlaGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly----- 875
QY 1766 CTCGGGGACTCCCTGCCCTGCATAAAAGAGGGTGTATTTCTACTGTTGTTATGTGTT 1825
Db 876 -----ProAspLysAlaProAlaLeuSerProGly-----SerCysCysPro----- 899
QY 1826 TGCTTTCAAATTGCTTAGTAGTACTCCTCAATTCAGTTATTATGAGCCAGCCTCAAGTTAG 1885
Db 890 -----ArgCysLeuProArgProAlaSerCysMetAlaPheGlyAspProHisTyrAr 907
QY 1886 AGAGCTAGGCTCT-----TCTTCAGGTGGACTCTGCCCAAATCACATACAAGTCAG 1936
Db 907 gThrPheAspGlyArgLeuLeuHisPheGlnGlySerCys-SerTyrValLeuAlaLysA 927
QY 1937 GTGCCATCAGGGTTTTTCCAGCCAGGCTGTGACAGGAGATATGGAGGGGGTCGG 1996
Db 927 spCysHisSerGlyAspPheSerValHis---ValThrAsnAspAspArgGlyArgSerG 946
QY 1997 GTTAGAGCTGG 2007
Db 946 lyValAlaIatrp 949
```

RESULT 10

US-10-189-971-24

; Sequence 24, Application US/10189971

; Publication No. US20030028907A1

; GENERAL INFORMATION:

; APPLICANT: Walker, D. Wade

; APPLICANT: Scoville, John

; APPLICANT: Turner, C. Alexander Jr.

; TITLE OF INVENTION: No. US20030028907A1el Human Kielln-like Proteins and Polynucleoti

; TITLE OF INVENTION: Same

; FILE REFERENCE: LEX-0360-USA

; CURRENT APPLICATION NUMBER: US/10/189, 971

; CURRENT FILING DATE: 2002-07-03

; PRIOR APPLICATION NUMBER: US 60/302,949

; PRIOR FILING DATE: 2001-07-03

; PRIOR APPLICATION NUMBER: US 60/315,634

; PRIOR FILING DATE: 2001-08-29

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 24

; LENGTH: 1342

; TYPE: PRT

; ORGANISM: homo sapiens

US-10-189-971-24

Alignment Scores:

Pred. No.:	0.0991	Length:	1342
Score:	124.00	Matches:	135
Percent Similarity:	28.32%	Conservative:	59
Best Local Similarity:	19.71%	Mismatches:	227
Query Match:	3.09%	Indels:	265
DB:	9	Gaps:	36

US-09-743-237-1 (1-2241) x US-10-189-971-24 (1-1342)

```
QY 558 GCCCTCTGGCT-----CAGGAATCTGTGTC----- 584
Db 420 AlaProCysAlaHisProArgGlnGlyProCysCysProSerCysAspGlyCysLeuTyr 439
QY 585 -----AAGTTCCTCATCATCCAGGAGGCA----- 608
Db 440 GlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAlaAlaCysHisLeu 459
```

```
QY 609 -----GAGGAGGCCTCCAGCTGCCCTCGGAAGAAAGAACTCCACGCCCATGGTG 656
Db 460 CysLeuCystripGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCys 479
QY 657 ATTGTGACGTAAAGAGGCGCCAGATGCTCTGCATAGACAACACTGTGCGCGAGGGAG 716
Db 480 ProPheProAlaArgGlyAsp-----CysCysProAspCysAspGlyCysGlu 495
QY 717 CTCAAAGCGCTCATCTGCTTCTCAGTAGCAGTACAGCAGCAGTCCCTCAGTCAGAG 776
Db 496 -----TyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspProArg 511
QY 777 CTCCTTAAGCAATGACAACCTTTAGTGGGAGACTTCTTGCCA----- 818
Db 512 GluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCysGlyArgArgProCys 531
QY 819 -----GTACCAGCG----- 827
Db 532 GluProProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGln 551
QY 828 -----AAGTTAAATCTCATCACAGGTGTAATGGAGCTCTCCCATCAGCTGCAAT 881
Db 552 GlyCysArgTyrHisGlyValThrThrAlaSerGlyGluThrLeuProAspProLeuAsp 571
QY 882 GGGGCTGCCTTT----- 899
Db 572 ProThrCysSerLeuCysThrCysGlnGluGlySerMetArgCysGlnLysLysProCys 591
QY 900 GGACCTGCTCTGCAAGGGCCACCCAAATAACTCTGTCTGGGTACTGTGACTGCTTCTCC 959
Db 592 AlaProAlaLeuCysProHisProSer-----ProGlyProCysPheCys----- 606
QY 960 AGCGGGGACTTCTGCAACAGCTGCAGCTGCAACAACCTGCGCCATGAGCTCGAGCGCTTC 1019
Db 607 -----ProValCysHisSerCysLeuSerGlnGlyArgGluHisGln----- 620
QY 1020 AAAGCCATAAAGCGTGTCTTTGATAGAAATCTCGAAGCTTTCACAAACAAAATGGGAAA 1079
Db 621 -----AspGlyGluGluPheGluGlyProAlaGlySer 631
QY 1080 -----GGCGTCTGGGAGCTCTAACTCGA----- 1106
Db 632 CysGluTrpCysArgCysGlnAlaGlyGlnValSerCysValArgLeuGlnCysProPro 651
QY 1107 -----CACAGCAAGGCTGCAACTGTAAAGCGC---TCAGGCTGC 1142
Db 652 LeuProCysLysLeuGlnValThrGluArgGlySerCysCysProArgCysArgGlyCys 671
QY 1143 CTGAAG----- 1148
Db 672 LeuAlaHisGlyGluGluHisProGluGlySerArgTyrValProProAspSerAlaCys 691
QY 1149 AACTACTGTGCTGCTATGAG-----GCCAAATCATGTGTTCTTCCATT 1193
Db 692 SerSerCysValCysHisGluGlyValValThrCysAlaArgIleGlnCysIleSerSer 711
QY 1194 TGCAAA-----TGCAATTCCTTGCACAAAACATGAGAA 1226
Db 712 CysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCysSerAspCysGluHis 731
QY 1227 AGTCCAGAACGAAAAATGCTGATGAGCACACCCCACTACATGAGCGCTGGGAGAC----TTT 1283
Db 732 GluGlyArgLys-----TyrGluProGlyGluSerPhe 742
QY 1284 GAGAGCAGCCATTATTTGTCCCGCCAGCCCAAGTTC-----TCAGGA 1322
Db 743 GlnProGly-----AlaAspProCysGluValCysIleCysGluProGlnProGluGly 760
QY 1323 CCTTCAAACTGAGAAAAATAGCGGCTTCTCTCTGTATCTCTCTGGGAGTAGTGAG 1382
Db 761 ProProSerLeuArgCysHisArgArg-----GlnCysProSer-----LeuValGly 776
QY 1383 GCCACATGTGCTGCTGTGCGCCCGCCAGGGGTGAGGAAGCAGCAGGAGGACTGTTCCCA 1442
```


RESULT 14

US-10-025-380-1116
: Sequence 1116, Application US/10025380
: Publication No. US20020182191A1

GENERAL INFORMATION:

: APPLICANT: Xu, Jiangchun
: APPLICANT: Lodes, Michael J.
: APPLICANT: Secrist, Heather
: APPLICANT: Benson, Darin R.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Stolk, John A.
: APPLICANT: Wang, Tongcong
: APPLICANT: Jiang, Yudi
: APPLICANT: Smith, Carole L.
: APPLICANT: King, Gordon E.
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Skeiky, Yasir A. W.
: APPLICANT: Fanger, Gary R.
: APPLICANT: Vedwick, Thomas S.
: APPLICANT: Carter, Darrick
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
: TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
: FILE REFERENCE: 210121.47IC14
: CURRENT APPLICATION NUMBER: US/10/025.380
: CURRENT FILING DATE: 2001-12-19
: NUMBER OF SEQ ID NOS: 1129
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1116
: LENGTH: 5405
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-025-380-1116

Alignment Scores:

Pred. No.: 0.368 Length: 5405
Score: 119.50 Matches: 122
Percent Similarity: 31.93% Conservative: 55
Best Local Similarity: 21.94% Mismatches: 162
Query Match: 2.95% Indels: 218
Gaps: 34

US-09-743-237-1 (1-2241) x US-10-025-380-1116 (1-5405)

```
QY 1479 TTCCAACTCCCTCCAGGATCATCTGCT-----CAGCCAGCTTGGGACAGT----- 1432
Db 1401 PheGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu 1420
QY 1431 -----GCTCCTGCTGCTTCTCCTCACCCTGGCCACAGGC----- 1396
Db 1421 GluValValProAsnSerProCysLeuProProThrProCysProProGlySerGluAsp 1440
QY 1395 -----AGGCACATGGGCTCCACTACTTCCAGGAGATACAGGAGAGGCTGCC 1345
Db 1441 CysIleProSerHisLysCysProGluLeuGluLysLysTrpGlnLysGlu----- 1458
QY 1344 TATTTTCTCAGTTTGGAGTCTCGACACTTGGCTGGGGACAAATAATGGCTGCTCT 1285
Db 1459 -----GluPheCysGlyLeuLeuSerSerProThrGlyProLeuSerSerCysHis 1475
QY 1284 CAA-----AGTCCCAGGCTCCA-----TGAGTGGG----- 1258
Db 1476 LysLeuValAspProGlnGlyProLeuLysAspCysIlePheAspLeuCysLeuGlyGly 1495
QY 1257 -----GTGTGCTCATCAGCATTTTTCGTTCTGGAGCTTCTTCATAGTTT 1213
Db 1496 GlyAsnLeuSerIleLeuCysSerAsn-----IleHisAlaTyr 1508
QY 1212 TCCAGCAATGC-----ATTTGCAATGGAAGACACATGATTGG 1171
Db 1509 ValSerAlaCysGlnAlaAlaGlyGlyHisValGluProTrpArgThrGluThrPheCys 1528
QY 1170 CCTCATGACTTCACAGTAGTTCTTCAGGACGCTTGAGGCTTACAGTTGACCCCTTGC 1111
```

```
Db 1529 ProMetGluCysProProAsnSerHisTyrGluLeuCysAlaAspThrCysSerLeuGly 1548
QY 1110 TGTGTGCAAGTT-----TAGCAGCTCCAGACGGCTT----- 1078
Db 1549 CysSerAlaLeuSerAlaProProGlnCysGlnAspGlyCysAlaGluGlyCysGlnCys 1568
QY 1077 ---TCCCATTTTGGTTGGAAAGCTTCAGGATTCTATCAAGACACGCTTTATGGCTT 1021
Db 1569 AspSerGlyPheLeuTyr----- 1574
QY 1020 TGAAGCGCTCGAGCTCATGGCGCAGTTGTTC-----AGCTGACGCTGTTGAGA 970
Db 1575 -----AsnGlyGlnAlaCysValProIleGlnGlnCysGlyCysTyrHis 1589
QY 969 AGTCCCGCTGGGAAGCAGTCACAGTACCCAGACAGAGTTATTTGGGTGGCCCTTGA 910
Db 1590 Asn----- 1590
QY 909 GAGCAGTCCAGAGGAAAGCGAGCCCATTCACAGCTGATGGGAGAGCTCCATTATCAA 850
Db 1591 -----GlyValTyrTyrGlu 1595
QY 849 CCTGTGTGATGAGATTTAACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 790
Db 1596 Pro-----GluGlnThrValLeu----- 1602
QY 789 TTGGCTTAGGGAGCTCTGACTGAGGAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 730
Db 1603 -----AspAsnCysArgGlnGlnCysThrCysHisAlaGly 1614
QY 729 GGAGCGCTTGGCTCCCTCGCGCCACAGTCT-----CTATGACAGAGCATCT 682
Db 1615 LysGlyMet-----ValCysGlnGluHisSerCysLysProGlyGlnValCysGlnProSer 1633
QY 681 GGGCGCTCTCTTTCA-----GCTGACAAATCACCATGGGCTGGAGTCTTCTTCCGAGGC 625
Db 1634 GlyGlyIleLeuSerCysValThrLysAspPro-----CysHisGly 1647
QY 624 ---AGCTGGAGGCTCTCT-----CTGCTCTCTGGGATGATGGAACTTGCACACAGG 577
Db 1648 ValThrCysArgProGlnGluThrCysLysGluGlnGlyGlyGlnGlyValCys----- 1665
QY 576 ATTCTGTAGCCAGAGGGGCTGGAAGCTGCTCTCAGGGCTGCGCTG----- 526
Db 1666 LeuProAsnTyrGluAlaThr-----CysTrpLeuTrpGlyAsp 1678
QY 525 GCACACTACCACTGCTTC-----TTTGATTTC-----AACACTTTTCTTGTCT 481
Db 1679 ProHisTyrHis-SerPheAspGlyArgLysPheAspPheGlnGlyThrCysAsnTyrVa 1698
QY 480 GCTTGTGCTGCTGGGAACTGGCAGTAGTTCATCGAGTCTCACCACAAATG-ATTTCCTCC 422
Db 1698 lLeuAlaThrThrGlyCys-----ProGlyValSerThrGlnGlyLeuThrPr 1714
QY 421 CTGTACACCTACCACTCAGGATGTAGACGACGAGAACTCAGAGTCTCTCTAGTTATAT 362
Db 1714 oPheThrValThrThrLysAsnGlnAsnArgGlyAsnProAlaVal-----SerTyrVa 1732
QY 361 AGCAATTCAGGAATTTGCTATATAGCTTAGCAATATG---GTTGAGAGACAGAAATTCAT 305
Db 1732 lArgValValThr-ValAlaAlaLeuGlyThrAsnIleSerIleHisLysAspGluIleG 1752
QY 304 CAAACTCTACATAGT-----TTAAGAAGTGTCCCTTATCTTACAGAGGGGAGAT 251
Db 1752 lLysValArgValAsnGlyValLeuThrAlaLeuProValSerValAlaAspGlyArgI 1772
QY 250 CATCAACTCAAAATCCTCAATATGGAAGTCTTTGGTGTGAAGAAAAAATCTCCC 191
Db 1772 leSerVal----- 1774
QY 190 AGCCCAACCCAACTACCAGGCTGAAGACTATCTGCAGAACACGCAAAAGGCTGGTT 131
```

Db 1775 -----ThrGlnGlyAlaSerLysAlaLeu----- 1782
QY 130 CTATCCCTAGCCCTGCTGGGGAGGTGGT-----GGAGTCA 92
Db 1783 --LeuValAlaAspPheGlyLeuGlnValSerTyrAspTrpAsnTrpArgValAspValt 1802
QY 91 CAAGTCCCAAGCGCCAGGAACACACACTGRTATCCCTTGTGGT 48
Db 1802 hrLeuProSerSerTyrHisGlyAlaValCysGlyLeuCysGly 1816
RESULT 15
US-09-922-217-1116
; Sequence 1116, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1116
; LENGTH: 5405
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1116
Alignment Scores:
Pred. No.: 0.368 Length: 5405
Score: 119.50 Matches: 122
Percent Similarity: 31.83% Conservative: 55
Best Local Similarity: 21.94% Mismatches: 162
Query Match: 2.95% Indels: 218
DB: 10 Gaps: 34
US-09-743-237-1 (1-2241) x US-09-922-217-1116 (1-5405)
QY 1479 TTCCAAACTCTCCAGGATCATCTGCT-----CAGCCAAAGCTGGGGAACAGT----- 1432
Db 1401 PheGlnLysProAsnGlySerGlnAlaGlyAsnAlaAsnGluPheGlyAsnSerTrpGlu 1420
QY 1431 -----GCTCCTGCTCTGCTTCTCCACCTCGCCGCGCAGCAGC----- 1396
Db 1421 GluValValProAspSerProCysLeuProThrProCysProProGlySerGluAsp 1440
QY 1395 -----AGGCACATGTCCTCCACTACTCCAGGAGATACAGGAAAGCCCTGCC 1345
Db 1441 CysIleProSerHisLysCysProProGluLeuGlyLysTyrGlnLysGlu----- 1458
QY 1344 TATTTTCTCAGTTTGGAGGCTCTCAGAACTTGGCTGGGCAATAATGCTGCTCT 1285
Db 1459 -----GluPheCysGlyLeuLeuSerSerProThrGlyProLeuSerSerCysHis 1475
QY 1284 CAA-----AGTCCCAAGCTCCA-----TGTAGTGGG----- 1258
Db 1476 LysLeuValAspProGlnGlyProLeuLysAspCysIlePheAspLeuCysLeuGlyGly 1495
QY 1257 -----GTCGCTCATACAGATTTTTCGTTCTGCGACTTCTCATAGTTT 1213
Db 1496 GlyAsnLeuSerIleLeuCysSerAsn-----IleHisAlaTyr 1508

QY 1212 TGCAAGCAATGC-----ATTGTCAAATGGAAGAACACATGATTTGG 1171
Db 1509 ValSerAlaCysGlnAlaAlaGlyGlyHisValCluProTrpArgThrGluThrPheCys 1528
QY 1170 CCTCATAGCACTCACAGTAGTCTTCTAGGAGCGCTGAGCGCTTACAGTTGCACCCCTTGGC 1111
Db 1529 ProMetGluCysProProAsnSerHisTyrGluLeuCysAlaAspThrCysSerLeuGly 1548
QY 1110 TGTGTGCAAGTT-----TAGCAGCTCCAGACGGCCCTT----- 1078
Db 1549 CysSerAlaLeuSerAlaProProGlnCysGlnAspGlyCysAlaGluGlyCysGlnCys 1568
QY 1077 ---TCCCATTTTGTGGTGGAAAGCTTCAGGATTTCTATCAAGACACGCTTTATGGCTT 1021
Db 1569 AspSerGlyPheLeuTyr----- 1574
QY 1020 TGAAGCGCTCGAGCTCATGCGCGCAGGTGTTC-----AGCTGCAGCTGTTGCAGA 970
Db 1575 -----AsnGlyGlnAlaCysValProIleGlnGlnCysGlyCysTyrHis 1589
QY 969 AGTCCCGCTGGAGAAGCAGTCACAGTACCAGACAGAGTTATTTGGGTGGCCCTTGCA 910
Db 1590 Asn----- 1590
QY 909 GAGCAGGCTCCAGAGGAAGGAGCCCAATTGACAGTGATGGGAGAGCTCCATTATCAA 850
Db 1591 -----GlyValTyrTyrGlu 1595
QY 849 CTTGTGTGATGAGATTTAACTTCGCTGGTACTGCGCAGAACTCTTCCACATAAGTTGTCA 790
Db 1596 Pro-----GluGlnThrValLeu----- 1602
QY 789 TTGCCTTAGGAGCTCTGACTGAGGGAACCTGCTCTGTCATCTACTGAGGAAGCAGAT 730
Db 1603 -----AspAsnCysArgGlnGlnCysThrCysHisAlaGly 1614
QY 729 GGAGCGCTTTGAGCTCCCTCGCGCCACAGTTGT-----CTATGCAGCAGCATCT 682
Db 1615 LysGlyMet---ValCysGlnGluHisSerCysLysProGlyGlnValCysGlnProSer 1633
QY 681 GGGCGCTCTCTTCA---GCTGACAAATCACCATGGGCGCTGGAGTCTTCTCCGAGGCG 625
Db 1634 GlyGlyIleLeuSerCysValThrLysAspPro-----CysHisGly 1647
QY 624 ---AGCTGGAGGCTCCT-----CTGCCCTCTGGGATGATGGAACTTGCACACAG 577
Db 1648 ValThrCysArgProGlnGluThrCysLysGluGlnGlyGlnGlyValCys----- 1665
QY 576 ATTCTTGAGCCAGAGGGGCGCTGGAAAGCTGCGTCTTCAGGGCTGCCCGCTG----- 526
Db 1666 LeuProAsnTyrGluAlaThr-----CysTrpLeuTrpGlyAsp 1678
QY 525 GCACATACACCTGCTTC-----TTTGATTTC---AACACTTTTCTTCTGCTT 481
Db 1679 ProHisTyrHis-SerPheAspGlyArgLysPheAspPheGlnGlyThrCysAsnTyrVa 1698
QY 480 GCTTGTGCTCTGGGAACCTGGCAGTATTCATCGGAGTCTCACCAAAATG-ATTTCGCC 422
Db 1698 lLeuAlaThrThrGlyCys-----ProGlyValSerThrGlnGlyLeuThrPr 1714
QY 421 CTGTACACCTTAGCAGCTAGGATGTAGAGGAGGAGATCAGAGTCTCTCTAAGTATAT 362
Db 1714 oPheThrValThrThrLysAsnGlnAsnArgGlyAsnProAlaVal-----SerTyrVa 1732
QY 361 AGCAATTTCAAGAAATGCTATATAGCTTAGCAATATG---GTTGAGAACAGAAATTCAT 305
Db 1732 lArgValValThr-ValAlaAlaLeuGlyThrAsnIleSerIleHisLysAspGluLeu 1752
QY 304 CAAACTCTACATAGT-----TTAAAGAACTCTCCCTTATCTTACAGAGGGGGAGAT 251
Db 1752 lLysValArgValAsnGlyValLeuThrAlaLeuProValSerValAlaAspGlyArgI 1772

```
QY 250 CATCAACTGAAAAATCCTCAATAATGGAAGTCTTTGGTGTGAAGAAAAAAATCTCC 191
Db 1772 leSerVal-----1774
QY 190 AGCCAAACCCAACTACCAGGCTGAAGACTATCTGCAGAACACGCAAAAGGCCCTGGGTT 131
Db 1775 -----ThrGlnGlyAlaSerLysAlaLeu----1782
QY 130 CTATCCCTAGCCCTGCTGGGGGAGGTGGT-----GGAGTCA 92
Db 1783 --LeuValAlaAspPheGlyLeuGlnValSerTyrAspTrpAsnTrpArgValAspValT 1802
QY 91 CAAGTCCCAACAGGCCAGGAACACACACTGTGTATCCCTTGTGGT 48
Db 1802 hrLeuProSerSertyrHisGlyAlaValCysGlyLeuCysGly 1816
```

Search completed: April 21, 2003, 11:48:24
Job time : 124.359 secs


```
QY 525 -----CCAGGCGCAGCCCTGAAGCGCAG----- 549
Db 42 oAlaGlnGlnAsnArgProGlyThrSerSer-MetIleGlnGlnHisAsnArgSerMetG 62
QY 550 --CTTTCAGGCGCCCTCGCTCAGGAATCCTGTGTGCAAGTCCCATCAT- 597
Db 62 luValasnGlnGlyLeuValylAspGluProIleAspThrSerSerHisArgValTyrV 82
QY 598 -----CCAGAGGCGCAGGAGGCGCTCCAGCTGCCCTCGGAGAAAGACCCAGCCC 649
Db 82 alProProProArgProValGlnArgLysPro-----GlyProSerThrPro--- 97
QY 650 CATGGCTATTGTTCAGCTGAAGAGGCGCGCCAGATGCTCTGCATAGACAACTGTGGCGC 709
Db 97 ----- 97
QY 710 GAGGAGCTCAAGCGCTCATCTGCTTCTCAGTACGATGACGAGCAGGATTTCCTCA 769
Db 98 --GlySerSerGlnTyrThrValArgAsnLeuSerAsnLeuSerGlySerProSerMet 117
QY 770 G-----TCAGAGCTCCCTAAG-----CCAATGACAACCTTGTAGTGGAA 807
Db 117 yrAspArgGlnProAlaSerLeuProArgThrValGlnProMetGlyLeuGluMetGlyA 137
QY 808 GACTTCTGCCAGTACCAGCGAAGTTAAATCTCATCAGAGTTGATATGGAGCTCTCC 867
Db 137 snSerGluGlnArgLysValTyrIleAspMetLysAspHisValSerHisIleArgLeuL 157
QY 868 CATCAGCTGTCAATGGGCTGCTTCCCTCTGAGCTGCTGCAAGGCGCCACCCAAA 927
Db 157 ysThrLysLysLysValPheAla-----ProGlyGlnArgLysProCysAsnC 173
QY 928 TAAC-----CTGTGCTGCTACTGTGACTGCTTCTCCAGCGGGAGCTTCTGCA 975
Db 173 ysThrLysSerGlnCysLeuLysLeuTyrCysAspCysPheAlaAsnGlyGluPheCysA 193
QY 976 ACAGCTGCAGCTGC-----AACAACTGCGCCATGAGCTCGAGCGCTTCAAG 1023
Db 193 rgAspCysAsnCysLysLysAspCysHisAsnIleGluTyrAspSerGlnArgSerLysA 213
QY 1024 CCATAAGCGCTGCTGTATAGAAATCCTGAAGCTTTTCAACCAAAATGGGAAAGGCC 1083
Db 213 laIleArgGlnSerLeuGluArgAsnProAsnAlaPheLysProLysIleGlyIleAlaA 233
QY 1084 GTCTGGAGCTGCTAACTTCGA-----CACAGCAAAAGGTGCAACTGTAAGCGCTAC 1137
Db 233 rgGlyGlyIleThrAspIleGluArgLeuHisGlnLysGlyCysHisCysLysLysSerG 253
QY 1138 GCTGCTGAAGAACTACTGTGAGTGTATCAGGCCAAATCATGTGTTCTTCTTCCATTGCA 1197
Db 253 lyCysLeuLysAsnTyrCysGlyCysTyrGluAlaLysValProCysThrAspArgCysL 273
QY 1198 AATGCATTGCTTGCAAAAAC-----TATGAAGAAAGT----- 1229
Db 273 ysCysLysGlyCysGlnAsnThrGluThrTyrArgMetThrArgTyrLysAsnSerGlyG 293
QY 1230 -----CCAGACGAAATTCGTATGAGC-----ACAC 1257
Db 293 lyAlaValSerAsnThrAsnAlaLeuMetSerLeuThrAsnAlaSerThrAlaThrP 313
QY 1258 CCCACTACATGAGCTGGGACTTTTGAGCAGCGCCATTAT-----TTGTCCCCAG 1308
Db 313 roAsp---SerGlyProGlySerValThrAspGluHisGlyAspAspTyrGluAspM 332
QY 1309 CCAAGTTCTCAGGACTCCAAACTGAGAAAAATAGCAGCGCTTCTCTGTATCTCT 1368
Db 332 etLeuLeuSerHisLysProLysValGluMetAspProArgPhe-----Prot 349
QY 1369 GG-----GAAGTAGTGGAGCCACATGCTGCTGCTGCGCCAGGGTG 1413
Db 349 rpTyrTyrMetThrAspGluValValGluAlaAlaThrMetCysMetValAlaGlnAlaG 369
QY 1414 AGGAACA-----GAGCAGGAGCACTGTTCCCAAGCTGGCTGAGC 1455
```

```
Db 369 luGluAlaLeuAsnTyrGluLysValGlnThrGluAspGluLysLeuIleAsnMetGluL 389
QY 1456 AGATGATCTCGAGGAGGTTTGGAAAGTGCTGTCGCAGAGATTCTC 1499
Db 389 ysLeuValLeuArgGluPheGlyArgCysLeuGluGlnMetIle 403
RESULT 2
B84585
hypothetical protein At2g20110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84585
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventre
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: B84420; MUID:20083487; PMID:10617197
A:Accession: B84585
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <STO>
A:Cross-references: GB:AE002093; NID:g4580462; PIDN:AAD24386.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g20110
A:Map position: 2
```

```
Alignment Scores:
Pred. No.: 2,89e-22 Length: 571
Score: 371.00 Matches: 92
Percent Similarity: 45.85% Conservative: 46
Best Local Similarity: 30.56% Mismatches: 101
Query Match: 9.26% Indels: 62
DB: 2 Gaps: 11
```

US-09-743-237-1 (1-2241) x B84585 (1-571)

```
QY 603 GAGGACAGGAGGAGCGCTCCAGCTGCCCTCGAAGAAAGACTCCAGGCCCATGGTATTGT 662
Db 3 GluGlyGluGluGlyAspLysPheProProLysThrAspGlu-----ValThrGln 19
QY 663 CAGCTGAAAGGAGGCGCCAGATGCTTCATAGACAACGTGTGGCGGAGGAGCTCAA 722
Db 20 GluSerMetLysSerAlaArgGlnLeuAspPheThrGly---GlySerSerAspValGlu 38
QY 723 GCGCTCCATCTGCTCTCCTCAGTACGATGACAGAGCAGTTTCCCTCAGTCAGAGCTCCCT 782
Db 39 HisSerHis-----SerAsnGlnAlaSerSerMetAlaAlaSerIlePro 54
QY 783 AAGCCAAATGACAACCTTACTGGGAAGACTTCTG----- 815
Db 55 SerProIleValThrValThrArgProIleIleThrSerGlnAlaProThrValAla 74
QY 816 -----CCAGTACCAGGAGTTAAATCTCATCATCACAGTTTCAT---AATGGA 860
Db 75 ThrProIleProProProProGlnSerGlnGlyIleIleLeuHisValProIleArgHis 94
QY 861 GCTCTCCCATCAGCTGTCAATGGGCTGCTTCCCTCTGGACCTGCTGCAAGGCCCA 920
Db 95 ProArgProGluSerProAsnSerMetProArgProAlaGlyGluThrArgAspGlyThr 114
QY 921 CCAAAA-----ATAACTCTCTCTGGGTACTGTGAC 950
Db 115 ProGlnLysLysLysGlnCysAsnCysLysHisSerArgCysLeuLysLeuTyrCysGlu 134
QY 951 TCGTTCTCCAGCGGGAGCTTCTGCAACAGCTGCAGCTGC-----AACACCTG 998
Db 135 CysPheAlaSerGlyThrTyrCysAspGlyCysAsnCysValAsnCysPheAsnVal 154
QY 999 CCGCATCTCGAGCGCTTCAAAGCCATAAAGCGGTGTCTGTATAGAAATCTCTGAAGCT 1058
Db 155 GluAsnGluProAlaArgGlnAlaValGluSerThrLeuGluArgAsnProAsnAla 174
```


Alignment Scores:

Pred. No.: 3,63e-12 Length: 658
Score: 248.50 Matches: 107
Percent Similarity: 39.95% Conservative: 50
Best Local Similarity: 27.23% Mismatches: 160
Query Match: 6.20% Indels: 76
DB: 2 Gaps: 18

US-09-743-237-1 (1-2241) x F71410 (1-658)

```
QY 441 GACTCCGATGACACTGCTCCCA---GGTTCACAGCGCAGCAAGCAAGAAAGTGT 497
Db 186 AspaSerGluLeuLeuIlePheAaspSerProAspAlaSerGluAlaPheArgCysPhe 205
QY 498 GAAATCAAGAAGCAGGTGGTGTGTGCCAGCGCGCAGCCCTGAAGACGAGCTTCCAG 557
Db 206 MetMetGlnArgAlaSerAsnSer-----GluAlaArgPheArg 218
QY 558 GCCCCTCTGGCTCAGGAATCCTGTTGCAAGTTCCTATCCAGGAGGAGGAGGCC 617
Db 219 AsnGlyValGluLysGlnThr---MetGlnHisAspSerAsnLysGluProGluSerAla 237
QY 618 TCCAGCTGCCCTCGGAAGAAGACTCCAGCCCATGTGATTTGTCAG-----665
Db 238 AsnAlaIleProTyrGluValAsnSerGly-----ValIleSerGlnAlaValSerLeu 255
QY 666 CTGAAGAGGAGCGGCCAGATGCTCTGCATAGAC-----698
Db 256 LeuHisArgGlyIleArgArgArgCysLeuAspPheGluMetProGlyAsnLysGlnThr 275
QY 699 -----AACTGTGGCGGAGGAGGAGCTCAAGCG-----725
Db 276 SerSerGluAsnAsnThrAlaAlaCysGluSerSerArgCysValValProSerIle 295
QY 726 ---CTCCATCTG-----CTTCCTCAGTACGATCACCAGACGATTTCCCTCAGTCA 773
Db 296 GlyLeuHisLeuAsnAlaIleLeuMetSerSerLysAspCysLysThrAsnValThrGln 315
QY 774 GAGCTCCTAAGCAATGACAACTTTAGTGGAGAGACTTCTGCCAGTACCAGCAAGTTA 833
Db 316 AspTyrSerCysSerAlaAsnIleGlnValGlyLeuGlnArgSerIleSerThrLeuGln 335
QY 834 AATCTCATACACAGTTGTATGATGAGCTCTCCCATCAGCTGTCAATGGGCTGCTTT 893
Db 336 AspSerLeuAspGlnThrGluAsnGluIleArgGluAspAlaAspGlnAspVal-----353
QY 894 CCCTCTGAGCTGCTGTGCA---GGGCCACCAAAATAACTCTCTCT 938
Db 354 ProValGluProAlaLeuGlnGluLeuAsnLeuSerSerProLysLys-----Lys 370
QY 939 GGTACTGTGACTGCTTCTCAGCGGGGACTTCTGC---AACAGCTGCAGCTGC-----989
Db 371 SerTyrCysGluCysPheAlaAlaGlyValTyrCysIleGluProCysSerCysIleAsp 390
QY 990 -----AACAACTGCCCATGAGCTCGAGGCTTCAAGCCATAAAGGCGTGTCTGTAT 1043
Db 391 CysPheAsnLysProIleHisGluAspValValLeuAlaThrArgLysGlnIleGluSer 410
QY 1044 AGAATCCTGAAGCTTTCACCAACCAAAATGGGAAGCGCTCTGGAGCT-----1094
Db 411 ArgAsnProLeuAlaPheAlaProLysValIleArgAsnSerAspSerValGlnGluThr 430
QY 1095 -----GCTAAACTTCGACACAGCAAGGTTGCAACTGTAAG 1130
Db 431 GlyAspAlaSerLysThrProAlaSerAlaArgHisLysArgGlyCysAsnCysLys 450
QY 1131 CGCTCAGCTGCCGTGAAGACTACTGTGATGCTATGAGGCCAAATATGTTCTTCC 1190
Db 451 LysSerAsnCysLeuLysLysTyrCysGluCysTyrGlnGlyValGlyCysSerIle 470
QY 1191 ATTTGCAATGCATGCTTCCAAACATATCAGAAAGTCCAGAACCAAAATGCTCATG 1250
Db 471 AsnCysArgCysGluGlyCysLysAsnAlaPheGlyArgLysAspGlySerSerIleAsp 490
```

```
QY 1251 AGCACACCCCACTACATGGAGCCTGGGGACTTTGAGAGACGCCATTATTTGCCAGCC 1310
Db 491 MetCluAlaGluGlnGluGlu-----GluAsnGluThrSerGluLysSerArgThrAla 508
QY 1311 AAGTTCTCAGAGCCTCCAAAACACTGAGAAAAAATAGGACGCGCTTCT-----1356
Db 509 Lys-SerGln-----GlnAsnThrGluValLeuMetArgLysAspMetSerSerAlaLe 526
QY 1357 CCTGTATCTCCCTGGGAAGTAGTGGAGGCCACATGTCCTGCTGGCCCGAGGTGAG 1415
Db 526 uproThrThrProThrProIleTyrArgProGluLeuValGlnLeu---ProPheSerSe 545
QY 1416 GAAGCAGACGAGCAGCAGCTGTTCCCAAGCTTGCGTG 1452
Db 545 rSerLysAsnArgMetProProGlnSerLeuLeu 557
```

RESULT 5

MEGF8 protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00209

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like moti

A:Reference number: Z14126; MUID:98360089; PMID:9693030

A:Accession: T00209

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1737 <NAB>

A:Cross-references: EMBL:AB011541; NID:G3449307; PIDN:BAA32469.1; PID:G3449308

A:Experimental source: brain; clone HG1392

C:Genetics:

A:Gene: MEGF8

A:Map position: 19q12

Alignment Scores:

Pred. No.: 0.00109 Length: 1737
Score: 146.00 Matches: 120
Percent Similarity: 30.86% Conservative: 38
Best Local Similarity: 23.44% Mismatches: 191
Query Match: 3.64% Indels: 163
DB: 2 Gaps: 30

US-09-743-237-1 (1-2241) x T00209 (1-1737)

```
QY 453 CTACTGCCAGTTCCTCCCAAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 509
Db 609 LeuLeuAlaProSerGlnGlyAlaLysArgAspArgMetArgAsnValArgGlySerSer 628
QY 510 GCAGGT---GGTAGTGTGCCAGCGCGCAGCCCTGAGACGCGCTTTCACAG-----557
Db 629 ArgGlyLeuGlyGlnValProGlyGluGlnProGlySerTrpGlyPheArgGluValArg 648
QY 558 -----GCCCTCTGGCTCAGGAATCTGTGTCAGAACTCTGTCAGAACTCCATCATCC 599
Db 649 LysLysMetAlaLeuTrpAlaLeuAla-----GlyThrGly 661
QY 600 CAGGAGCAGCAGGAGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCTC-----650
Db 662 GlyPheLeuGluIleLeuSer-----ProHisLeuLysGluProArgProArgLeuPhe 679
QY 651 -----ATGGTGATTTGTCAGCTGAAAGAGGCGCGCCAG 683
Db 680 HisAlaSerAlaLeuLeuGlyAspThrMetValValLeu-----GlyGlyArgSer 696
QY 684 -----ATGCTCTGCATAGACAACTGTGCGCGAGGAG 716
Db 697 AspProAspGluPheSerSerAspValLeuLeuTyrGlnValAsnCysAsnAla-----714
QY 717 CTAAAGCGCTCCTCCTGCTCCTCAGTACGATGACAGCAGCAGTTCCTCCTCAGTCAG 776
```

```
Db 715 -----TrpLeuLeuProAspLeuThrArgSerAlaSer----- 725
QY 777 CTCCTTAAGCCAAATAGTGGGAGACTTCTCCAGTACCACCGAAGTTAAAT 836
Db 726 ValGlyProMetGluSerValAlaHisAlaValAlaValGlySerArgLeu 745
QY 837 CTCATCACAGGTTGATTAATAGACTCTCCCATCAGCTGTCAATGGGCTCCCTTTCC 896
Db 746 TyrIleSerGlyGlyPheGlyValAlaLeuGlyArgLeuLeuAlaLeuThrLeuPro 765
QY 897 TCTGGACCTGCT-----CTGCAAGGGCCACCCAAATTAACCTCTGCTGGGTACTGTGAC 950
Db 766 ProAspProCysArgLeuLeuSerSerProGluAlaCysAsnGlnSerGlyAlaCysThr 785
QY 951 ----TGCTTCTCC-----AGCGGGGAC-----TTCTGCAACAGCTGC 983
Db 786 TrpCysHisGlyAlaCysLeuSerGlyAspGlnAlaHisArgLeuGlyCysGlyGlySer 805
QY 984 AGCTGCAACACCTG-----CGCCATGAGCTCGAGCGCTTCAAAGCCATAAGGCG 1034
Db 806 ProCysSerProMetProArgSerProGluGluCysArgArgLeuArgThrCysSerGlu 825
QY 1035 TGTCTTGATAGAAATCTCAAGCTTTCACCAAAATGGGAAAGCGCTGTGGAGCT 1094
Db 826 CysLeuAlaArgHisProArgThrLeuGlnProGlyAspGlyGlu----- 840
QY 1095 GCTAAACTTCGACACAGCAAGGTCG---AACTGAAGCGCTCAGGCTGCCTG----- 1145
Db 841 AlaSerThrProArgCysLysTrpCysThrAsnCysProGluGlyAlaCysIleGlyArg 860
QY 1146 -----AAGAACTACTGT----- 1157
Db 861 AsnGlySerCysThrSerGlnAsnAspCysArgIleAsnGlnArgGluValPheThrPala 880
QY 1158 ----GAGTGTATGAGGCCAAATC-----ATGTGTCTTCCATTTCG 1196
Db 881 GlyAsnCysSerGluAlaAlaCysGlyAlaAlaAspCysGluGlnCysThrArgGluGly 900
QY 1197 AAATGATTCGTCGAAAACTATGAAGAACTCCAGAA---CGAAATGCTGATGACG 1253
Db 901 LysCysMetTrpThrArgGlnPheLysArgThrGlyGluThrArgArgIleLeuSerVal 920
QY 1254 ACACCCCACTAC-----ATGAGCGCTGGGGAC 1280
Db 921 GlnProThrTyrAspTrpThrCysPheSerHisSerLeuLeuAsnValSerProMetPro 940
QY 1281 TTTGAGACGACCATATTTGTCGCCAGCAAGTCTCAGGACCTCCAAACTGAGAAA 1340
Db 941 ValGluSerSer-----ProLeuProCysProThrProCysHisLeuLeuPro 957
QY 1341 AATAGCAGGCGCTTCTCTGTATCTCTGGGAAGTAGTGAGGCCACATGTCCTGCCGTG 1400
Db 958 Asn-----CysThrSer-----CysLeu 963
QY 1401 CTGGCCAGGCTGAGGACGAGGAGCACTGT----- 1436
Db 964 AspSerLysGlyAlaAspGlyGlyTrpGlnHisCysValTrpSerSerSerLeuGlnGln 983
QY 1437 -----TCCCAAGCTTGCTGAGCAGATGATCTGGAGGAGTTTGGAAAGTGCTGTGCG 1490
Db 984 CysLeuSerProSer-TyrLeu-----ProLeuArgCysMetAlaGlyGlyCysGln 1000
QY 1491 CAGATTCTCCACATCGAGTTCAAGTCCAAAGGGCTCAAAATGAGTAGCTGCAAGCTGG 1550
Db 1000 yArgLeu-----LeuArgGlyProGluSerCysSerLeuGlyCysAlaGlnAlaPh 1017
QY 1551 TAAA-----GGGAATGCCTGTGGCAAGCGCTCAGC 1580
Db 1017 rGlnCysAlaLeuCysLeuArgArgProHisCysGlyTrpCysAlaTrpGlyGlnAs 1037
QY 1581 CCTGGGAATCTGCACGAGGAGCTGGTCCCGAGGGA-----GG 1619
Db 1037 pGlyGlyArgCysMetGluGlyLeuSerGlyProArgAspGlyLeuThrCysGln 1057
```

```
QY 1620 AGCAGAGGCGCGCATCATGGCCAGGTCAAGTCTGCT 1653
Db 1057 yArgProGlyAlaSerTrpAlaPheLeuSerCys 1068

RESULT 6
T08852
Lustrin A - California red abalone
C:Species: Haliotis rufescens (California red abalone)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08852
R:Shen, X.; Belcher, A.M.; Hansma, P.K.; Stucky, G.D.; Morse, D.E.
J. Biol. Chem. 272, 32472-32481, 1997
A:Title: Molecular cloning and characterization of lustrin A, a matrix protein from S.
A:Reference number: Z16496; MUID:98070424; PMID:9405458
A:Accession: T08852
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1428 <SHE>
A:Cross-references: EMBL:AF023459; NID:q2723361; PIDN:AAB95154.1; PID:q2723362
A:Experimental source: tissue type mantle (shell and pearl nacre); cell type pallial
C:Superfamily: antileukoproteinase repeat homology
C:Keywords: extracellular matrix; extracellular protein
F:1382-1426/Domain: antileukoproteinase repeat homology <ALP>
```

Alignment Scores:

Pred. No.:	0.0169	Length:	1428
Score:	131.50	Matches:	141
Percent Similarity:	29.36%	Conservative:	51
Best Local Similarity:	21.56%	Mismatches:	244
Query Match:	3.28%	Indels:	218
DB:	2	Gaps:	31

US-09-743-237-1 (1-2241) x T08852 (1-1428)

```
QY 498 GAAATCAAGAAGCAGGTGGTAGTGTG-----CCA 527
Db 297 GluIleSerCysAlaGlyGlyAlaCysProValAsnThrValCysValAlaHisPro 316
QY 528 GCGGCGAGCCCTGAGACGACGCTTTCAGGCGCCCTCTGGCTCAGAAATCTGTGCAAG 587
Db 317 SerGlyAlaProAlaValCysCysPheLysPro-----AlaGlyProThrThrProGln 334
QY 588 TTCCCATCATCCAGAGGCGAGGAGGCTCCAGCTGCCCTCGGAAGAAGACTCCAGC 647
Db 335 ProProThrIleProGlnProProThrThrProSerSerProThrGlyAspProCysGlu 354
QY 648 CCATGGTGATTTGTGAGCTGAAAGGAGGCGCCAGATGCTCTGCATAGAC----- 698
Db 355 ProGlyValAsnValAsnCysThrAlaGlyThrCysArgLeuValValAspCysArgPhe 374
QY 699 ----AACTGTGGCGGAGGAGCTCAAAGCGCTCCATCTGCTCCTCAGTACGATGACCAAG 755
Db 375 ProGlyCysProAla-----ValProLysCysValAspPro 386
QY 756 AGCAGTTTCCCTCAGTCAGAGCTCCCTAAGCCAATGACAACTTTAGTGGGAAGACTTCTG 815
Db 387 SerSerLysProSer-----LeuAsnCysSerIleGlyAsp----- 398
QY 816 CCAGTACCAGCG-----AAGTTAAATCTCATCACAGGTTGATAATGGAGCTCTC 866
Db 399 -----ProAlaLeuAsnProAsnLeuGlnGluIleSerCysValGlyAlaAlaCys 416
QY 867 CCATCAGCTGTCAATGGGGCTGCCCTTCCCTCTGGA---CCTGCTCTG----- 911
Db 417 ProArgAsnThrAlaCysPheAlaProSerGlySerProAlaValCysCysThr 436
QY 912 CAAGGCGCACCCAAATAACTCTGTCTGGGTACTGTGACTGCTCTCCAGCGGGGACTTC 971
Db 437 SerGlyProProArgProGluProProSerProSerPro-----ProThrGlyAspPro 454
QY 972 TGCAAC-----AGCTGACAGTGCACAAACACTGCGC-----CAT 1004
Db 972 TGCAAC-----AGCTGACAGTGCACAAACACTGCGC-----CAT 1004
```


Qy	744	ACTGAGAACGACATGGAGCGCTTTT	GAGCTCCCTCCGCCACACAGTTGTCTATGCAGACA	689
Db	624	ThrAlaSerProProGly-----	-----ProProProGlyrLysArgAlaProSer	639
Qy	684	TCTGGCGCCTTCCTTTACGCTGACAAATCACCATGGGCTGGAGTCTTTTCTTCCGAGGCG	625	
Db	640	ProGlyAlaTyrlYsthrAla-----	-----ThrProProGlyrLyrlYsthrProGlySerProPro	657
Qy	624	AGCTGGAGGCGCTCCTCGCTCCTCGGATGATGGGAACATTGCAACAGGATTCTCTGAGCCA	565	
Db	658	SerPheargThrGlyThrProProGlyrTyrlArgGlyThr-----	-----SerProPro	673
Qy	564	GAGGGGCTCGAAGTGCGCTTCACGGGTCGCG-----	-----CCTGGCACACATACCAC	514
Db	674	AlaGlyProGlyThrPheLys--	--ProGlySerProThrValGlyProGlyProLeuPro	693
Qy	513	CTGCT	509	
Db	693	roAla	694	
RESULT 8				
T03455				
ALR protein - human				
C:Species: Homo sapiens (man)				
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000				
C:Accession: T03455				
P:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapala				
Oncogene 15, 549-560, 1997				
A>Title: Structure and expression pattern of human ALR, a novel gene with s				
A:Reference number: Z14954; MUID:97388474; PMID:9247308				
A:Status: preliminary;				
A:Molecule type: mRNA				
A:Residues: 1-4957 <PRA>				
A:Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2355				
C:Genetics:				
A:Gene: ALR				
A:Map position: 12				
C:Superfamily: human ALR protein				
C:Keywords: alternative splicing				
Alignment Scores:				
Pred. No.:	0.0846	Length:	4957	
Score:	123.50	Matches:	127	
Percent Similarity:	34.99%	Conservative:	56	
Best Local Similarity:	24.28%	Mismatches:	189	
Query Match:	3.05%	Indels:	152	
DB:	2	Gaps:	26	
US-09-743-237-1 (1-2241) x T03455 (1-4957)				
Qy	1816	CAACAAGTAGAAAATCACCTCTTTATGCAGGGCAGGAGTCCCAGAG---	GGCCTC	1766
Db	3378	GlnGlnGlnGlnGlnGlnGlnPhelGlnGlnGlnGlnGlnGlnMetGlyLeu	3399	
Qy	1759	CCCAAGGCAGAGGAGGCGCAGTGTCCATCCCCAGGGTGGCTCTCGCTGTGCATCTTAGG	1700	
Db	3398	LeuAsnGlnSerArgThrLeuLeuSerProGlnGlnGlnGln-----	Gln	3411
Qy	1699	ATACCTTGAGTAGCTGGCCAGTACCATGCAGATCATCTCAGACCTCACAGCTGACCTGGC	164	
Db	3414	GlnValAlaLeuGlyProGlyMetProAlaLysProLeuGlnHisPheSerProGly	3433	
Qy	1639	CATGATCGCGGCGCTCTGCTCTCCCTCCCTGGCCACCACTCTCTCTCGGTGCAGATTCACGAG	158	
Db	3434	AlaLeuGlyProThrLeuLeuLeuThrGlyLysGluGlnAsnThr-ValAsp-----	3455	
Qy	1579	CTGAGGCTTGCCACAGGCATTCCTCCCTTACCACTTGCAGCTTACTCAATTTTACGCCCC	152	
Db	3451	-----ProAlaValSerSerGluAlaThrGlu-----	GlyPr	3466
Qy	1519	TTGACTTGAACTCGATGTGGGAATCTCGCACAGCACCC-----	TTCANAAC	147

Db	3451	oSerThr	---HisGlnGlyProLeuAlaIleGlyThrThrProGlnUserMetaIaTh	3480
Qy	1471	TCCTCCAGGATCATCTGCTCAGCAACAGCTTGGGAACAGTGTCTCTGCTCTGCTT	-----	1417
Db	3480	rGluProGly	---GluValIysProSerLeuSerGlyAspSerGlnLeuLeuLeuValG1	3499
Qy	1416	-----CCTCACCCCTGGGCCAGCAGGACACATGTGGCTCCACTTA	---CTTCCACG	1367
Db	3499	nProGlnProGlnProGlnProSerSerLeuGlnLeuGlnProProLeuArgLeuProG1	3519	
Qy	1366	-GAGATACAGSAGAAGCCCTGCTATTTTCTCAGTTTTGGAGGTCTCTGAGAACTTGGC	1308	
Db	3519	yGlnGlnGlnGlnValSerLeuLeuHisThrAlaGlyGlySerHisGlyGlnLe	3539	
Qy	1307	TGGSGACAAATAATGGCTGCTCTCAAGTCCCGAGCTCCATGTAGTGGGTGTGCTCAT	1248	
Db	3539	uGly	-----	3540
Qy	1247	CAGCATTTTTCGTTCTGGACTTTCTTCATAGATTTTGTCAAGCAATGCATTTGCAATGGA	1188	
Db	3541	-----SerGlySerSerSerGluAlaSerSerValProHisLeuLeuAlaG1	3556	
Qy	1187	AGAACACATGATTTTGGCCTCATGACATCAGCTAGTCTTTCAGGCAGCCTGAGCCGTT	1128	
Db	3556	nProSerValSerLeuGlyAsp	-----GlnProGlySerMe	3568
Qy	1127	ACAGTTGTCAC	-----CCTTTGCTGTGCGAAGTTTA	1097
Db	3568	tThrGlnAsnLeuLeuGlyProGlnGlnProMetLeuGluArgProMetGlnAsnAsnTh	3588	
Qy	1096	---GCAGCTCCAGACGGCCTTCCCATTTT	-----GGTTGGAAAGCTTC	1053
Db	3588	rGlyProGlnProProLysProGlyProValLeuGlnSerGlyGlnGlyLeuProGlyVa	3608	
Qy	1052	AGCATTTCTATCA	-----	1040
Db	3608	lGlyIleMetProThrValIyGlnLeuArgAlaGlnLeuGlnGlyValLeuAlaLysAs	3628	
Qy	1039	-----AGACAGCGCTTTATGCTTTTGAAGCGCTCGAGCTCATGGCGCAGGTGTT	990	
Db	3628	nProGlnLeuArgHis	-----LeuSerProGlnGlnGlnGlnLeuGlnAlaLeuLe	3646
Qy	989	GCAGCTGCAGCTGTTTGCAGAAGTCCCGCTCGAGAAGCAGTCA	-----	947
Db	3646	uMetGlnArgGlnLeuGlnGlnSerGlnAlaValArgGlnThrProProThrGlnGluPr	3666	
Qy	946	-----CAGTACCCAGACAGATTTATTTTGGTGGCCCTTCAGACAGCT	-----	902
Db	3666	oGlyThrGlnThrSerProLeuGlnGlyLeuLeuGlyGlyCysGlnProGlnLeuGlyPh	3686	
Qy	901	-CCAGAGGAAAGGACGCCCATTCACAGCTGATGGGAGAGCTCCATTATCAACCTGTGT	843	
Db	3686	eProGlyProGlnThrGlyProLeuGlnGluLeuGly	-----	3698
Qy	842	GATGAGATTAACTTCGCTGGTACTGGCAGAAGTCTTCCCACTAAAGTTGTCATTGGCTT	783	
Db	3699	-----AlaGlyProArgProGlnGlyProPro-ArgLeuProAlaProp	3713	
Qy	782	AGGAGCTCTGACTCAGGAAACTGCTCTGTT	-----CATCGTACTGAGGAAGCAGATG	729
Db	3713	roGlyAlaLeuSerThrGlyProValLeuGlyProValHisProThrProProProSerS	3733	
Qy	728	GAGCGTTTGGAGCTCCCTCGCGCCACAGTTGCTCTATCCAGGACATCGGGCGCTCCTTT	669	
Db	3733	erProGln	---GluProLysArgProSerGlnLeu	3743
Qy	668	CAGCTGCACAAATCACCATGGGGCTGGAGTCTTCTCCGAGGCGAGCTGGAGCCCTCCTC	609	
Db	3744	-----ProSerPro	-----SerSerGlnLeuProThrGluAlaGlnL	3756
Qy	608	TGCCTCCCTGGGATGATGGAACTTGCACAGGATTTCTGAGCCAGAGGGCCCTGGAAAGC	549	
Db	3756	euProProThrHisProGlyThr	-----ProLysProGlnGlyPro	3769

```

QY 548 TGGCGTCTTCAGGCGCTGGCAGCTACACACCTGCTCTCT-----TTGATTTCAAC 495
Db 3770 --ThrLeuGluPro--ProProGlyArgValSerProAlaAlaAlaGlnLeuAlaAspTh 3788
QY 494 ACTTTTCTTCGTGCTGCTGCTTTGGGAACCTGGCAGTAGTTCATCGGAGTCTCACCA 435
Db 3788 rIeuPheSerLysGlyLeuGlyProTrpAspProAlaGlnLeuAlaGlnGlnGly 3808
QY 434 A 434
Db 3808 s 3808

RESULT 9
T03454
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03454
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli,
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with stro
A:Reference number: Z14954; MUID:97388474; PMID:92477308
A:Accession: T03454
A>Status: Preliminary; translated from GR/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-5262 <PRA>
A:Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

Alignment Scores:
Pred. No.: 0.085 Length: 5262
Score: 123.50 Matches: 127
Percent Similarity: 34.99% Conservative: 56
Best Local Similarity: 24.28% Mismatches: 189
Query Match: 3.05% Indels: 152
DB: 2 Gaps: 26

US-09-743-237-1 (1-2241) x T03454 (1-5262)

```

Qy	1816	CAACAAGTAGAAAATCACCCTCTTTTATGCAGGCGAGGAGTCCCCAGAG---GGCCTC	17650
		::: ::: :::	
Db	3683	GlnGlnGlnGlnGlnGlnPheGlnGlnGlnGlnGlnGlnGlnMetGlyLeu	3702
Qy	1759	CCCAGGACACAGGAGGCCAGTGCATCCACAGGCTGGCTGCTGCCTGCACCTTTAGG	17000
		:::	
Db	3703	LeuAsnGlnSerArgThrLeuLeuSerProGlnGlnGlnGlnGln---Gln	3718
Qy	1699	ATACCTTGAGTAGCTGGCCAGTACCATGCAGATCACTCAGACCTTCACAGCTGCACCTGGC	1640
		::: ::: ::: :::	
Db	3719	GlnValAlaLeuGlyProGlyMetProAlaLysProLeuGlnHisPheSerSerProGly	3738
Qy	1639	CATGATGCGCGGCGCTCTGCTCTCCCTGGGACACAGCTTCTCGGTGCAGATGCCACGG	1580
		:::	
Db	3739	AlaLeuGlyProThrLeuLeuLeuThrGlyLysGluGlnAsnThr-ValAsp-----	3755
Qy	1579	CTGAGGCTGCCACACAGCATTCCTCCCTTACACAGCTTGACGCTACTCANTTTCAGCCCC	1520
		:::	
Db	3756	-----ProAlaValSerSerGluAlaThrGlu-----GlyPr	3766
Qy	1519	TTGGACTTGAACTCGATCGAGATCTGCGACAGGCACC-----TTCCAAC	1472
		:::	
Db	3766	oSerThr---HisGlnGlyGlyProLeuAlaTiedGlyThrProGluSerMetAlaThr	3785
Qy	1471	TCCTCAGGATCATCTGCTACGCCAAGCTTGGGAACAGTCTCCTGCTCGTGT---	1417
		::: :::	
Db	3785	rGluProGly---GlnValLysProSerGlyAspSerGlnLeuLeuValGln	3804
Qy	1416	-----CCTCACCTGGGCCACAGGACGATGGCGCTCCACTA---CTTCCAG	1367


```

Db 476 aglyAsnProSerLeuHisGly- - - - - SerProIleLeuSerLeuAs 491
QY 1588 TTCCAGGGCTGAGGCTTGCACAGGATCCCTTTTACCAGCTTGACGCTACTCAATT 1529
Db 491 nserGln- - - - - ProHisVal- - - - - ArgAsnGlyAspAsnI 502
QY 1528 TTCAGCCCTTGAGCTTGAACATCGATGCGAGAACTCGACAGGACCTTC- - - - 1477
Db 502 eSerAlaPro- - - - - AsnValLysAlaSerGluSerGlyHisPheAlaSerTh 518
QY 1476 - - - - - CAACCTCTCCAGGATCA- - - - - TCTGCTCAGCAAGCTTGGGAA 1436
Db 518 rArgGlnAsnLysProGlnProProAlaSerIleSerValPro- - - - - 534
QY 1435 CAGTCTCTCTGCTCTCTCACCCTGGCCAGGAGGAGGACATGTGGCTCCACT 1376
Db 535 - - - - - AlaProAlaPheIle- - - - - TyrProAlaAsnHisLeuGlnProValMe 550
QY 1375 ACTTCCAGGAGATACAGGAGGCTGCTATTTTCTCAGTTTGGAGGCTCTGAG 1316
Db 550 tValProSerLys- - - - - SerSerArgPro- - - - - 558
QY 1315 AACTTGGCTGGGCAATAATAGCTGCTCTCAAAGTCCCAGGCTCCATGTAGTGGGT 1256
Db 559 - - - - - ThrLysSerProHisLeu- - - - - AlaValGlyLe 568
QY 1255 GTGCTCATCAGCAATTTTCTGCTCGGACTTCTTCATAGTTTTCGAAGCAATCATTTG 1196
Db 568 uAlaSerAlaAsnPhe- - - - - 573
QY 1195 CAATGGAAGAACACATGATTTTGGCCTCATAGCACTCACAGTAGTCTTTCAGGACG- - 1138
Db 574 - - - - - SerHisProSerSerSerAlaSerG 582
QY 1137 - - - - - CTGAGCGCTTACAGTTGACGCTTTCAGGCTGCTG 1109
Db 582 uValSerSerProTyrLeuThrValIleProAsnAlaTyrSerPheGlnLeuSerSe 602
QY 1108 TGTGGAAGTTTAGCAGCTCCAGGAGGCTTTCCTCCATTTTGGTGTGGAAGCTT- - - 1054
Db 602 rThiLeaArgGlyGlyThrProSerGlnAlaValProPheTyrAsnGlySerPheTyrSe 622
QY 1053 - CAGGATTTCTATCAAGACAGCCTTTATGCTT- - - - - TGAAGCGCTCGAGC 1007
Db 622 rProGlnMetPheGlnGlnProGlnProGlnProLeuGlnArgGlnSerGlnAlaGlnArgG 642
QY 1006 TCATGGCGCAGGTTGTCAGCTGCAGCTGTGCAAGAGTCCCGCTGGAGACAGCTCA 947
Db 642 uSerLysAlaSerSerCysSerSerSerHisArgGlnPro- - - - - 656
QY 946 CAGTACCCAGACAGAGTTATTTTGGTGGCCCTTGCAGAGCAGGTCCAGAGGAAAGGCA 887
Db 657 - - - - - GlnValSerValAsnSerLeuSerSerGlnAlaAsnValGlnGlnHisArgG 674
QY 886 GCCCATTTGACAGCTGATGGGAGAGCTCCATATFCAACCTGTGTGATCAGATTTAACTTC 827
Db 674 n- - - - - 674
QY 826 CTGGTACTGGCACAAGCTTCCCACTAAAGTTGTCATTTGCTTAGGAGCTCTGACTGA 767
Db 675 - - - - - MetSerGlnLysPheGluValAla- - - - - GlyValAsnThrAs 687
QY 766 GGGAAACTGCTCTGCTCATCTACTAGGAGAGCAGATGGAGCGCTTTGAGCTCCCTCCG 707
Db 687 pSerArg- - - - - GlySerHisThrGlnLysGlyGly- - - - - 697
QY 706 CCACAGTTGCTATGCAGAGCATCTGGCGCCTCTCTTTCAGCTGCACAAATCACCATGGG 647
Db 698 - - - - - ProPheGlyGlnIleMetAlaAlaProValGlnProGlnAs 711
QY 646 CTGGAGTCTTCT- - - - - TCCGAGGCGCAGCTGGAGGCTCTCT- - - - - 610

```

```

Db 711 nPheSerMetSerPheAlaSerIleAlaSerSerAlaProProAlaThrLeuAsnPheSe 731
QY 609 - - - - - CTGCTCTCTGGGATG- - - - - 592
Db 731 rSerAsnGlyTyrHisIleSerThrProProGlyValAlaHisGlnLysAsnHisGlnSe 751
QY 591 - - - - - GGAACCTGCAACAGGATTCCTGAGCCAGGAGGCGCTGCTTCTTTGATT 500
Db 751 rSerGluAlaLysThrGlyGlySerCysSer- - - - - SerAsnAlaGluAspPro- - 768
QY 553 AAAGCTCGCTCTCAGGCG- - - - - TCCTGCTGGCAGCTACACCTGCTGCTTTGATT 500
Db 769 - LysLysAsnLeuGlnGlyLysProGlnGlyMetMetAsnGlyHisThrLeuVal- PheA 788
QY 499 TCAACACTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 440
Db 788 spAsnProSerArgThrLeuAsnPheValSerGlyThrTipPro- - - - - 802
QY 439 CACCAAAATGATTTCCTGCTGTA- - - - - 417
Db 803 - - - - - ProProAlaAlaThrAlaIleAsnGlyAspProSerValPheThrG 818
QY 416 - CACCTAGCAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 359
Db 818 lnHisLeuThrGlnArg- - - - - GlnGlnGlnSerGlyArgSerLysMetMetMetThrH 836
QY 358 AATTTCAAGCAATGCTATAGCTTACGATATGCTTACAGAGGGGGAGATCATCACTGAAA 299
Db 836 isSerGlnAlaAspSerValSerAlaThrSerSerGlnThrLysAsnProAlaThrSer- 855
QY 298 TCTACATAGGTTTAAAGAACTGTCCCTTATCTTACAGAGGGGGAGATCATCACTGAAA 239
Db 856 - - - - - SerSerLeuThrSerCysThrSerLeuAsnLeuLysGlnPheGlnSerGlnG 874
QY 238 AATCTCAATATGGAAGCTTCTTGTGTGTCGAAGAAAAAATCTCCAGCCCAACCCAA 179
Db 874 lnIleArgThrHisGlyGlnThrGlnIleSerPheAlaAlaProThrAsnProGlnPro 894
QY 178 ACTACAGGCTGAAGACTATCTGCAAGAACAGCAA- - - - - AAGGCCCTGGTCTATCCCTA 122
Db 894 er- - - - - GlnGlyLysGlnGlyArgSerGlyGlySerSerPro 907
QY 121 GC- - - - - CCTGCTGGGCGGAGGTGGTGGAGTCAACA 89
Db 907 erValThrGlySerAlaSerHisGlyLysProAlaAsnSerLysValSerAsnSerLysA 927
QY 88 GT- - - - - CCAACAGCCAGGAAACACACA 66
Db 927 laLeuLeuLeuSerProValProLeuSerGlnGluHisThr 940
RESULT 11
S50832
atrophin-1 - human
C:Species: Homo sapiens (man)
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
C:Accession: S50832
R:Nagafuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.;
Nature Genet. 8, 177-181, 1994
A:Title: Structure and expression of the gene responsible for the triplet repeat disco
A:Reference number: S50832; MUID:95144175; PMID:7842016
A:Accession: S50832.
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1184 <NAG>
A:Cross-references: EMBL:D31840
C:Genetics:
A:Gene: GDB:DRPLA; B37
A:Cross-references: GDB:270336; OMIM:125370
A:Map position: 12p13.31-12p13.3112p-12p
Alignment Scores:
Pred. No.: 0.162 Length: 1184
Score: 119.50 Matches: 125

```


Percent Similarity:	29.91%	Conservative:	44
Best Local Similarity:	22.12%	Mismatches:	206
Query Match:	2.95%	Indels:	191
DB:	2	Gaps:	24

US-09-743-237-1 (1-2241) x S50832 (1-1184)

QY		1991	C C C C T C C A A T A T C T C G T C - - - - - A C A G C C T G G C C T G G A A A A C C C C T G A	1944
Db		254	P r o P h e t H r T h r P r o F l e S e r V a l S e r S e r G l y A l a S e r G l y A l a P r o P r O t h r	273
QY		1943	T G C C A C C T G A C T T G T A T G T G A T T T G G C A G A G T C C A C T G A A C A G A G C C T A G C T C T C T	1884
Db		274	L y s P r o P r o T h r T h r P r o v a l - G l y G l y A s n L e u P r o S e r A l a P r o P r o P r o A l a --	292
QY		1883	A A C T T G A G G C T G G C T C A T A N T A A C T T G A A T G A A G G T A C T A T A A C A A T T G A A G C A A A	1824
Db		292	- - - - -	292
QY		1823	C A C A T A A C A A G A G A A A A T C A C C T C T T T T A T G C A G G C A G G A G T C C C C A G A G G	1764
Db		293	-- -- Asn Phe Pro His Val Thr Pro Asn Leu Pro ProproProAl	306
QY		1763	C T C C C C A G G A C A G A G G C C A G T G T C C A T C C C C A G G T G G C T C T C G C T T G C A C T T	1704
Db		306	a L eu A r g p r o L e u A s n A s n A l a S e r A l a S e r P r o G l y L e u G l y A l a - - - - -	322
QY		1703	T A G G A T A C C T T G A G T A G G C T G C C A G T A C C A T G C A G A T C A C T C A G A C C T C A C A C C	1649
Db		323	-- -- Gl n P r o L e u P r o G l y H i s L e u P r o S e r P r o f r y A l a M e t G I	336
QY		1648	-- -- T G A C C T G G C A T G A T G C G C G G C C T G C T G C T C C T C C C T G G G - -	1610
Db		336	y G l n G l y M e t G l y L e u P r o P r o G l y P r o G l u L y s G l y P r o T h r L e u A l a P r o S e r P r	356
QY		1609	- C A C A G C T T C T C G G T G C A G A T T C C A G G C T G A G G C T T C C C A C A G G C A T T C C C T T W A	1551
Db		356	O H i s S e r L e u P r o - P r o A l a S e r S e r A l a P r o A l a P r o P r o M e t A r g P h e P r o f r y S	376
QY		1550	C A G C T T G C A C G C T A C T C A A T T T T C A G C C C T T G A C T T G A A C T G A T C T G G A G A A P C T G	1491
Db		376	e r S e r - - - - - SerSerSerSerSerAala	384
QY		1490	G C A G A G C A C T T C C A A A C T C C C A G G A T C A T C G T C A C C A A G C T T G G G A A C A G T G	1431
Db		384	I a l a S e r S e r S e r S e r S e r S e r S e r A l a S e r P r o P h e P r o A l a S e r G l n A	404
QY		1430	C T C C T C C T C T C T C T C A C C C T G G C C A G C A G G C A C A T G T G C C C T C C A C T A C T T C	1371
Db		404	I a L e u P r o S e r T y r P r o H i s S e r P h e P r o - - - - - P r o P r o T h r S e r L	418
QY		1370	C C A G G A G A T A C A G G A G A G G C C T A T T T T T T C A G T T T T G G A G G T C C T G A G A A C T T	1311
Db		418	E u S e r V a l S e r A s n G l n P r o P o l y S t y r T h r G l n P r o S e r L e u P r o S e r G l n A l a V a l T	438
QY		1310	G G - - - - - C T G G G A C A A A A T A A T G G C T G C T	1287
Db		438	r p S e r G l n G l y P r o P r o P r o P r o P r o P r o P r o T y r G l y A r g L e u L e u A l a A s n S e r A s n A l a H	458
QY		1286	C T C A A A G T C C C A G G C C C A T G A T A G T G G G G T G T C A T C A C A G C A T T T T G C T T G G A C T	1227
Db		458	I s P r o G l y P r o P r o S e r T h r G l y A l a G l n S e r T h r A l a H i s P r o P r o V a l S e r T	478
QY		1226	T T C T T C A T A G T T T T G C A A G C A A T C A T T T G C A A T T G G A A G A A C - - - - -	1183
Db		478	H r H i s H i s H i s H i s G l n G l n G l n G l n G l n G l n G l n G l n G l n G l n H	498
QY		1182	- - - - - A C A T G A T T T T G G C C C T C A T - - - - - A	1164
Db		498	I s H i s G l y A s n S e r G l y P r o P r o P r o G l y A l a P h e P r o H i s P r o L e u G l u G l y c l y S	518
QY		1163	G C A C T A C - - - - - A G T A G T T C T T C A G G C A G C C T G A G C G C T T T A C A	1125

Db	518	erSerHisHisAlaHisProTyAlaMetSerProSerLeuGlySerLeuArgProTyP	538
Qy	1124	GTTC-----CACCCCTTGCTGTGCAAGTTTACGAGCTC	1090
Db	538	roProGlyProAlaHisLeuProProHisSerGlnValSerTySerGlnAlaGlyP	558
Qy	1089	CCAGACGGCCTTTCCCACTTTTGGTTGGAAGCTTCAGGATTTCATCAAGACACACGCT	1030
Db	558	roAsnGlyPro--ProVal-----	563
Qy	1029	TTATGGCTTTGAAGCCCTCGAGCTCATGCCGAGGTGTTCAGCTGCAGCTGTTCGAGA	970
Db	564	-----SerSerSerSer-----AsnS	569
Qy	969	AGTCCCGCTGGAGAAGCAGCTACAGCTACCCACAGAGATTATTTGGTGGCTTCGCTTCG-	911
Db	569	erSerSerSerThrSerGlnGlySerTy-----ProCysS	581
Qy	910	-----AGAGCAGCTCCAGAGGAAAGGCAGCCCCATTGACAGCTGATGGGA	865
Db	581	erHisProSerProSerGlnGlyProGlnGlyAlaProTyProPhe-----	596
Qy	864	GAGCTCCATTATCAACTGTGTGATCAGATTAACTTCGCTGGTACTGGCAGAACTCTTC	805
Db	597	--ProProValProThr-Val-----ThrThrSerSerAlaThrLeuSerThrValLe	613
Qy	804	CCACTAAAGTTGTCATTGGCTTAGGGAGCTCTGACTGAGGGAACTGCTGTGGTCATCGT	745
Db	614	-----AlaThrValAlaSerSerProAlaGlyTyLys	624
Qy	744	ACTGAGGAAGCAGATGGAGCGCTTTGAGCTCCCTCGCCACAGTGTCTATGCAGAGCA	685
Db	625	ThrAlaSerProGly-----ProProTyGlyLysArgAlaProSer	640
Qy	684	TCTGGGGCCTCTTCAGCTGCACAATCACCATCGGGCTGGAGTCTTCTTCGAGAGGC	625
Db	641	ProGlyAlaTyrlsThrAla-----ThrProProGlyTyLysProGlySerProPro	658
Qy	624	AGCTGGAGGCCTCTCGCTCTCGGATGATGGGAACCTTGCACAGGATTCCCTGAGCCA	565
Db	659	SerPheargThrGlyThrProProGlyTyArgGlyThr-----SerProPro	674
Qy	564	GAGGGCTCGGAAGCTGCTCTTCAGGCTGCCG-----CCTGGCACACTACCAC	514
Db	675	AlaGlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuProp	694
Qy	513	CTGCT	509
Db	694	roAla	695

RESULT 12

KR301
S09257

homeotic protein Hox A4 - chicken

N;Alternate names: homeotic protein Chox1-4

C/Species: Gallus gallus (chicken)

```
C:\Species: gallus gallus (chicken)
C:\Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 21-Jul-2000
```

C;Accession: S09257; S10883

R; Sasaki, H.; Yokoyama, E.; Kuroiwa, A.

Nucleic Acids Res. 18, 1739-1747, 1990

A:Title: Specific DNA binding of the tw

A:Reference number: S09256: MUID:90245562: PMID:19708666

A: Accession: S09257

A: Accession: 309237
A: Molecule type: mRNA

A: MOLECULE TYPE: mRNA
A: Residues: 1-309 <SAS>

A;RESIDUES: I-309 <SAS>
A:CROSS-references: EMBL.: X52670. NID.: a63218. PTD.:CAA36896 1. PTD.:a63219

A;Cross-References: EMBL:X52670; NID:g6321
R:Scotting P J : Hewitt M : Keynes R J

R;scotting, P.J.; Hewitt, M.; Key
Nucleic Acids Res 18 3090 1990

Nucleic Acids Res. 18, 3999, 1990

A;Title: Isolation and analysis of chick homeobox cDNA

A; Reference number:

A;Accession: S10883

A; Molecule type: mRNA

A;Residues: 207-273 <SCO>

C;Function:
A;Description: control of embryonic development by tissue- and stage-specific regulation
C;Superfamily: homeotic protein Hox D4; homeobox homology
C;Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation
F;210-266/Domain: homeobox homology <Hox>

Alignment Scores:			
Pred. No.:	0.214	Length:	309
Score:	117.50	Matches:	58
Percent Similarity:	36.3%	Conservative:	14
Best Local Similarity:	29.2%	Mismatches:	81
Query Match:	2.90%	Indels:	45
DB:	1	Gaps:	9

US-09-743-237-1 (1-2241) x S09257 (1-309)

1829	QY	AGCAAAACATACACACAGT	AGAAAATCA	CCCTCTTTTTATGCAGGCAGGAGTCCCC	1770	
24	Db	ThrGlnHisSerGlySerAla	GlySerAla	GlySerAla	GlyHisProHisProHisPro	43
1769	QY	AGAGGCGCTCCCAAGGAGG	CGCCAGTCTCAT	CCCCAGGGTGGCTCGCTGCTT	1710	
44	Db	HisAlaProPro	-----Pro	ProProProProHisLe	57	
1709	QY	GCACCTTAGGATACCTT	GAGTAGGCTGGCCAGTAC	CATGCAGATCACTCAGACCTC	1650	
57	Db	uHisAlaAlaHisPro	-----Gly	ProAlaLeuProGluTyrPhe	ProArgProArg	75
1649	QY	CTGACCTGCCATGATGG	CGGCGCTGCTCCTCC	TGGGCACAGCTTCTCGGTGCAG	1590	
75	Db	gGluProGlyTyrGlnAla	---ProAla	AlaProProGlyProPro	92	
1589	QY	ATTCCAGCGGCTGAGGCT	TGCCACAGGCA	TTCCTTACAGCTTCGACGCCTACTCAAT	1530	
93	Db	----ProGluAlaLeuTyr	ProAlaGlnAlaProSerTyr	ProGlnAlaProTyrSer-T	111	
1529	QY	TTTCAGCCCTTGGACTT	GAACTCGATGTGGAGAA	TCTGGCAGCAGGCCTTCCAACTC	1470	
111	Db	YrSerSer	-----Ala	GlySerAlaAlaProGlyProGlu	Prop	125
1469	QY	CTCCAGGATCATCTCT	CAGCCCAAGCTTGGG	AACAGTGTCTCTGCTCTCTCC	TCAACC	1410
125	Db	roProGlyAlaSerPro	ProPro-----Pro	ProAlaLeuGlyHisP	140	
1409	QY	CTGGGCCACGAGGCG	CACATGTGGCTC	CACTACTTCCC-----AGGA	1365	
140	Db	roGlyProAla-----	GlnPro	LeuLeuProGlyHisAlaLeu	GlnArg	155
1364	QY	GATACAGGAGAAGCG	CTCGCTTATTTTCT	CAGTTTGGAGGCTCTCAGAACTTGCTGTG	1305	
155	Db	rycsgluAlaAlaPro	Ala-----Ala	GlyAlaArg	165	
1304	QY	GGACAAAATANGGCT	CTCTCTCAAGTCCC	AGGCTCCATCTAGTGGGTG	1255	
165	Db	lyThrGlyProGlyCys	AlaLeuLeuProAsp	YrSerLeuProGlyLeu	181	

RESULT 13

T13954

MEGF6 protein - rat

C; Species: *Rattus norvegicus* (Norway rat)

C;Species: *Rattus norvegicus* (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text change 21-Jul-2000

C;Accession: T13954

R; Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34,

A;Title: Identification of high-molecular-weight proteins with multiple ECF-like motifs

A; Reference number: Z14126; MUID:98360089; PMID:9693030

A; Accession: T13954

A; Status: preliminary; translated from GB/EMBL/DDB.

A; Molecule type: tRNA

A: Molecule type: mRNA
A: Residues: 1-1574 <NAK>

A;RESTQUES: I-13/4 <NAK>
A:Cross-references: EMBL:AR011532. NID:33449293. PIDN:BA32463 1. PID:33449294

A: Experimental source: strain Sprague-Dawley; brain
cross-references: EMBL:AB011532; NID:g3449293; PI

A; Experiment
C: Genetics:

A;Gene: MEGF6

Alignment Scores:

Pred. No.:	0.242	Length:	1574
Score:	117.50	Matches:	152
Percent Similarity:	28.6%	Conservative:	50
Best Local Similarity:	21.5%	Mismatches:	228
Query Match:	2.93%	Indels:	275
DB:	2	Gaps:	41

US-09-743-237-1 (1-2241) x T13954 (1-1574)

[illegible]

Db 1117 AspLysCysGlnSerSerCysValSerGlyThrPheGlyValHis 1131
QY 1377 GTGGAGGCACATGTGCTCGCTCGCTGGCC 1406
Db 1132 CysGluGluHisCysAlaCysArgLysGlyAlaSerCysHisValThrGlyAlaCys 1151
QY 1407 CAGGGTGAAGAGCAGACGAGGAGCTGTTCGCCAAGCTTG 1448
Db 1152 PheCysProGlyThrArgGlyProHisCysGluGlnAla---Cys-ProArgGlyTr 1170
QY 1449 GCTG-----AGCAGATGATCTGGAGAGTTT 1475
Db 1170 PheGlyGluAlaCysAlaGlnArgCysLeuCysProThrAsnAlaSerCysHisVal 1190
QY 1476 GGAAGTGCCTGTCGACAGATCTCCACATCGAGTTCAAGTCCAAAGGGCTGAATGAG 1535
Db 1190 IThrGlyGluCysArg-----CysProGlyPheThrGlyLeuSe 1204
QY 1536 TAGCCTGCAAGCT-----GGTAAA----- 1554
Db 1204 rCysGluGlnAlaCysGlnProGlyThrPheGlyLysAspCysGluHisLeuCysGlnCy 1224
QY 1555 ----GGGAATGCCCTGGCAAGCTCAGCCCTGGATCTCCACCGAGGAGCTGTGTC 1610
Db 1224 sProGlyGluThrTrpAlaCysAspProAlaSerGlyValCysThrCysAlaAlaGlyTy 1244
QY 1611 CCAGGGA-----GGAGCAGAGCCCGCATCATGCCAGGTCAGCTGTGA 1655
Db 1244 rHisGlyThrGlyCysLeuGlnArgCysProSerGlyArgTyGlyPro---GlyCysGl 1263
QY 1656 GGTCTGAGTGTCTGCAT-----GGTACTGGCCAGCCTACTCAAGGTATCTTAAAGTG 1708
Db 1263 uHisIleCysLysCysLeuAsnGlyGlyThrCysAspProAlaThrGlyAlaCysTyCy 1283
QY 1709 CAAGCAGGACAGCCACCTGGGGATGGACACTGCCCTCTCTGCTGGGAGGCCCTC 1768
Db 1283 s-----ProAlaGlyPheLeuGlyAlaAspCysSerLeuAlaCysProGl 1298
QY 1769 TGGGACTCCCTGCTGCATATAAAGAGGGTGATTTCTACTTGTGTGTATGTTGTC 1828
Db 1298 nGlyArgPheGlyPro-----SerCysAlaHisValCysAl 1310
QY 1829 TTTCAAAATGCTTAGTAGT----- 1847
Db 1310 aCysArgGlnGlyAlaAlaCysAspProValSerGlyAlaCysIleCysSerProGlyLy 1330
QY 1848 -ACCTCCATTCAAGTTATTATGAGCCAGCCTCAAGTTAGA----- 1886
Db 1330 sThrGlyValArgCysGluHisGlyCysProGlnAspArgPheGlyLysGlyCysGluLe 1350
QY 1887 -GAGCTAGGCTTCTTTCAGGTGGACTGCG---CCAAATCATACATACAGTCAGTGCC 1942
Db 1350 uLysCysAlaCysArgAsnGlyGlyLeuCysHisAlaThrAsnGlySerCysSerCysPr 1370
QY 1943 ATCAGGGGTTTTTCAGGCCAGGCGCTGTGACAGG-----AGATATGGGAG 1987
Db 1370 OleuGly---TrpMetGlyProHisCysGluHisAlaCysProAlaGlyArgTyGlyAl 1389
QY 1988 GGGGTGCGGTGAGAGTGGGTTTCTGGATTTTTCGTTTTCCTTCTCTGCTATTTC 2047
Db 1389 aAlaCysLeuLeuGluCys-----PheCysGlnAsnAsnGlySerCys----- 1403
QY 2048 TGCTTGAAGTGAGAAACT-----TGCTCTCTGTCACACCTTTTCTCCATAATTAC 2098
Db 1404 -----GluProThrThrGlyAlaCysLeu-Cys-GlyProGlyPheTyGlyGln 1419
QY 2099 TGCTGCAGC----- 2107
Db 1420 AlaCysGluHisSerCysProSerGlyPheHisGlyProGlyCysGlnArgValCysGlu 1439
QY 2108 -----GTGCGCTGTGACCAAGTCACAGTGAACCTCAGACACCAGAGAGGTGAGGT 2155
||| |||||

Db 1440 CysGlnGlnGlyAlaProCysAspProValSer-----Gly 1451
QY 2156 GCCTTATTATGCCCA-----CACITTTGTTGTTTGT-----TGTCAGATATAA 2197
Db 1452 GlnCysLeuCysProAlaGlyPheHisGlyGlnPheCysGluLysGlyCysGluSerGly 1471
QY 2198 CCTTTC 2203
Db 1472 SerPhe 1473
RESULT 14
S08167
Balbiani ring 3 protein - midge (Chironomus tentans)
C:Species: Chironomus tentans
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 21-Jul-2000
C:Accession: S08167
R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.
J. Mol. Biol. 211, 331-349, 1990
A:Title: The balbiani ring 3 gene in Chironomus tentans has a diverged repetitive str
A:Reference number: S08167; MUID:90172404; PMID:1689777
A:Accession: S08167
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1700 <PAU>
A:Cross-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058
C:Genetics:
A:Gene: BR3
A:Map position: 4
C:Superfamily: unassigned Balbiani ring proteins
Alignment Scores:
Pred. No.: 0.244 Length: 1700
Score: 117.50 Matches: 100
Percent Similarity: 31.11% Conservative: 49
Best Local Similarity: 20.88% Mismatches: 147
Query Match: 2.93% Indels: 183
DB: 2 Gaps: 29
US-09-743-237-1 (1-2241) x S08167 (1-1700)
QY 579 TGTTCGAAGTTCCCA---TCATCCAGGAGCAGAGGAGGCTCCAGC----- 623
Db 152 CysCysGlyCysProValAsnMetGlnGluProAlaAspGlyCysThrLysProLeuLe 171
QY 624 -----TCCCTCCGAAGAAAGAC----- 641
Db 172 TrpAspLysValAspCysArgCysGluCysProLeuLysLysAspCysGlyLysAsnArg 191
QY 642 -----TCCAGCCCCATGGTGTGTTGTCAGCTGAAGGA-----GCGGCC 680
Db 192 AspTrpSerAspSerSerCysSerCysGluCysLysGlyAspGlyLysCysGlnGlySer 211
QY 681 CAGATGCTCTGCATAGACAACACTGTGGCGGAGGAGCTCAAAGCGCTCATCTGCTTCT 740
Db 212 LysIleTrpCysLysAsnAsnCys--ArgCysIleCysProThrAlaGluProAlaGlyG 231
QY 741 CAGTACGATGACACAGCAGATTCCCTCAGTCAGAGCTCCCTAAGCCATGACAACTTTA 800
Db 231 LysCysSerAlaProLeuLysTrpAsp-----AspAspLysCysS 244
QY 801 GTGGGAAGACTTCTCCAGTACCCAGCGAAGTTAAATCTCAT-----CACACAGGTTGAT 854
Db 244 erCysAlaCysProAlaLysMetGluGluLysLysGluLysCysValGluSerGlyLysI 264
QY 855 AATGAGCTCTCCATCAGCTGTCAATGGGCTGCTTCCCTCTGGAGCTCTGCTGCA 914
Db 264 leTrp---AsnProAsnThrCysGluCysGlyCysAlaGlnLeu---AsnCysProAspA 282
QY 915 GGGCCACCCAAATAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
Db 282 snLysLysAlaAsnLysGlu-ThrCysGlnCysGluLysLysGluValLysLysCysAsn 301
QY 960 AGCGGGGAC---TTCTGC---AACAGCTGCAGCTGCAACAACTCGCCATGAGCTCGAG 1013

A: Molecule type: protein
A: Residues: 'D', 21-23/45-47; 147-148; 150-152 <XIA>
R: Yang, S.X.; Pollock, H.G.; Rawitch, A.B.
Arch. Biochem. Biophys. 327, 61-70, 1996
A: Title: Glycosylation in human thyroglobulin: location of the N-linked oligosaccharide
A: Reference number: S62778; MUID: 96201348; PMID: 8615697
A: Accession: S62778
A: Molecule type: protein
A: Residues: 69-84, 196-210, 'T', 212-213; 476-492; 523-539; 741-770; 811-848; 850-953; 938-950; 126-1783; 'D', 1785, 'R': 1999-2017; 2241-2258; 2270-2285, 'P', 2287-2307; 2578-2584 <YAN>
R: Ielari, T.; Cochaux, P.; Targovnik, H.M.; Suzuki, M.; Shimoda, S.; Perret, J.; Vassart, J. Clin. Invest. 88, 1901-1905, 1991
A: Title: A 3' splice site mutation in the thyroglobulin gene responsible for congenital
A: Reference number: 155565; MUID: 92091498; PMID: 1752952
A: Accession: 155565
A: Status: translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 90-91; 160-161 <TEI>
A: Cross-references: GB: S71821; NID: g240935; PIDN: AAB20665.1; PID: g240936
A: Note: mutant splice form
C: Comment: The thyroglobulin molecule is produced in the thyroid gland and is the precursor of thyroglobulin.
C: Genetics:
A: Gene: GDB-TG
A: Cross-references: GDB: 120434; OMIM: 188450
A: Map position: 8q24.2-8q24.3
A: Introns: 23/1; 59/2; 92/1; 160/1; 213/2; 249/1; 297/1; 359/1; 726/1; 921/1; 1000/1 #st
C: Complex: homodimer
C: Function:
A: Description: precursor of thyroid hormones thyroxine and triiodothyronine; iodine store
C: Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat homology
C: Keywords: alternative splicing; duplication; glycoprotein; homodimer; iodine; thyroid
F: 1-19/Domain: signal sequence #status predicted <SIG>
F: 20-2767/Product: thyroglobulin #status predicted <MAT>
F: 34-92/Domain: thyroglobulin type I repeat homology <THY1>
F: 96-160/Domain: thyroglobulin type I repeat homology <THY2>
F: 168-237/Domain: thyroglobulin type I repeat homology #status atypical <THY3>
F: 301-358/Domain: thyroglobulin type I repeat homology <THY4>
F: 608-658/Domain: thyroglobulin type I repeat homology <THY5>
F: 662-726/Domain: thyroglobulin type I repeat homology <THY6>
F: 730-921/Domain: thyroglobulin type I repeat homology #status atypical <THY7>
F: 923-1072/Domain: thyroglobulin type I repeat homology #status atypical <THY8>
F: 1076-1144/Domain: thyroglobulin type I repeat homology <THY9>
F: 1148-1209/Domain: thyroglobulin type I repeat homology <THY10>
F: 1455-1468/Region: type II repeat
F: 1469-1485/Region: type II repeat
F: 1486-1502/Region: type II repeat
F: 1602-1722/Region: type IIIa repeat
F: 1723-1891/Region: type IIIa repeat
F: 1892-1994/Region: type IIIb repeat
F: 1995-2128/Region: type IIIb repeat
F: 2129-2186/Region: type IIIa repeat
F: 2227-2725/Domain: cholinesterase homology <ACE>
F: 24-2572, 2586/Modified site: thyroxine (Tyr) #status predicted
F: 76, 198, 484, 529, 748, 816, 947, 1219, 1348, 1364, 1715, 1773, 2012, 2249, 2294, 2591/Binding site:
F: 110, 436, 1868, 2121/Binding site: carbohydrate (Asn) (covalent) #status absent
F: 149/Modified site: dehydroalanine (Tyr) #status predicted
F: 2765/Modified site: triiodothyronine (Tyr) #status predicted

Alignment Scores:
Pred. No.: 0.306 Length: 2767
Score: 116.50 Matches: 144
Percent Similarity: 32.42% Conservativity: 80
Best Local Similarity: 20.84% Mismatches: 269
Query Match: 2.91% Indels: 199
DB: 1 Gaps: 35

US-09-743-237-1 (1-2241) x UIHU (1-2767)

Qy 447 GATGAACACTGCCAGGTTCCCAAGGCGCAGCAAGCAAGAAAGTTTGAAATCAA 506
Db 1132 GUGLULeuArgProGlySerSerSerAlaGlnCysProSerLeuCysAsnValLeu 1151
Qy 507 GAAGCAGGT-----GGTAGTGCCAGCGGCGCAGCCCTGAA 542

Db 1152 LysSerGlyValLeuSerArgValSerProGlyTyrValProAlaCysArgAlaGlu 1171
Qy 543 GACGAGCTTTC-----CAGGCCCTCTGGCTCAGGAATCCTGT---TGCAGATTC 590
Db 1172 AspGlyGlyPheSerProValGlnCysAspGlnAlaGlnGlySerCysTrpCysValMet 1191
Qy 591 CCATCATCCAGGAG-----GCAGAGGAGGCC 617
Db 1192 AspSerGlyGluGluValProGlyThrArgValThrGlyGlnProAlaCysGluSer 1211
Qy 618 TCCAGCTCCCTCGGAAGAACACTCCAGCCCATGGTGTATTGTGACGCTGAAGAGGAGC 677
Db 1212 ProArgCysProLeuProPheAsnAlaSerGluValVal-----Gly 1225
Qy 678 GCCAGATGCTCTGC-----ATAGACAACACTGTGGCGGAGGAGCTCAACGCGTCCAT 731
Db 1226 GlyThrIleLeuCysGluThrIleSerGlyProThrGlySerAlaMetGlnCysGln 1245
Qy 732 CTGCTTCTCAGTACGATGACCCAGAGCAGTTTCCCT----- 767
Db 1246 LeuLeuCysArgGlnGlySerTrpSerValPheProGlyProLeuIleCysSerLeu 1265
Qy 768 -----CAGTCAGAGCTCCCTAAGCCCAATGACAACCTTACTGGGAAGACTT 812
Db 1266 GluSerGlyArgTrpGluSerGlnLeuProGlnProArgAlaCysGlnArgProGlnLeu 1285
Qy 813 ---CTGCCAGTACACGCAAGTTAATCTCATCATCACAGGTTGATANTGGAGCTCTCCA 869
Db 1286 TrpGlnThrIleGlnThrGlnGlyHisPheGlnLeuGlnLeuProGlyLysMetCys 1305
Qy 870 TCAGCTGTCAATGGGGCTGCCCTTCCCTCTGCACCTGCTCTCAAGGGGCCCAAAATA 929
Db 1306 SerAlaAspTyrAlaGlyLeuLeuGlnThrPheGlnValPheIle---LeuAspGluLeu 1324
Qy 930 ACTCTGTCTGGTACTGT-----GACTGCTTCTCCAGCGGGGACTCTTCGCAAGCTGC 983
Db 1325 ThrAlaArgGlyPheCysGlnIleGlnValLysThrPheGlyThrLeuValSerIlePro 1344
Qy 984 ACCTCCACACCTCCGC----- 1001
Db 1345 ValCysAsnAsnSerSerValGlnValGlyCysLeuThrArgGluLeuGlyValAsn 1364
Qy 1002 -----CATGAG 1007
Db 1365 ValThrTrpLysSerArgLeuGluAspIleProValAlaSerLeuProAspLeuHisAsp 1384
Qy 1008 CTCGAGCGCTTCAAGCCATAAAGCGGTGCTTTGATAGAAATCCTGAA----- 1055
Db 1385 IleGluArgAlaLeuValGlyLysAspLeuLeuGlyArgPheThrAspLeuIleGlnSer 1404
Qy 1056 ---GCTTTCCACCAAAATGGGAAGCCGCTCTGGGAGCTGCTAAACTTCGACACAGC 1112
Db 1405 GlySerPheGlnLeuHisLeuAspSerLysThrPheProAlaGluThrIleArgPheLeu 1424
Qy 1113 AAAGGCTGCAACTGT-----AAGCGCTCAGGCTGCTGAGAACTAC--- 1154
Db 1425 GlnGlyAspHisPheGlyThrSerProArgThrArgPheGlyCysSerGluGlyPheTyr 1444
Qy 1155 -----TGTGAGTGTATGAGGCC 1172
Db 1445 GlnValLeuThrSerGluAlaSerGlnAspGlyLeuGlyCysValLysCysHisGluGly 1464
Qy 1173 AAAATCATGTGTTCTTCCATTGCAANTGCATTGCTGCAAA-----AACTATGAAGAA 1226
Db 1465 SerTyr-----SerGlnAspGluGluCysIleProCysProValGlyPheTyrGlnGlu 1482
Qy 1227 AGTCCAGAAGCAAAATGCTGATGAGCACACCC-----CACTACATGAGCGCTGG 1277
Db 1483 GlnAlaGlySerLeuAlaCysValProCysProValGlyArgThrThrIleSerAlaGly 1502
Qy 1278 GACTTTGAGAGCGCATTTATTGTCCCA----- 1307

Db 1503 AlapheSerGlnThrHisCysValThrAspCysGlnArgAsnGluAlaGlyLeuGlnCys 1522
QY 1308 -----GCCAAGTTCTCAGGACCTCCAAAATGAGAAAAATAGCAGCGCTTCTCC 1358
Db 1523 AspGlnAsnGlyGlnTyrArgAlaSerGlnLysAspArgGlySerGlyLysAlaPhe--- 1541
QY 1359 TGTATCTCC-----TGGAGTAGTCTGGAGGCC-----ACA 1388
Db 1542 CysValAspGlyGluGlyArgLeuProTyrTrpGluThrGluAlaProLeuGluAsp 1561
QY 1389 TGTGCTGCTGCTGGCCAGGCTGAGAA----- 1418
Db 1562 SerGlnCysLeuMetMetGlnLysPheGluLysValProGluSerLysValIlePheAsp 1581
QY 1419 -----GCAGAGCAGGAGCACTGTTCCCAAGCTTGGCTGAGCAGATGATC 1463
Db 1582 AlaAsnAlaProValAla-ValArgSerLysValProAspSerGluPhePro-----Va 1599
QY 1464 CTGGAGGAGTTTG-----GAAGTGCTGTCGCGAGATTCACACATCGAG 1508
Db 1599 lMetGlnCysLeuThrAspCysThrGluAspGluAlaCysSerPhePheThrValSerTh 1619
QY 1509 TTCAAGTCCAAAGGGCTGAAATGAGTACGCTGCAAGCTGTAAGGGGAATGC----- 1563
Db 1619 rThrGluPro-----GluIleSerCysAspPh 1628
QY 1564 ----CTGTGCAAGCCTCAGCCCTGGGAATCTGCAC-----GAGGAAGCTGGTGCCCA 1613
Db 1628 eTyrAlaTrpThrSerAspAsnValAlaCysMetThrSerAspGlnLysArgAspAlaLe 1648
QY 1614 GCGAGGACGAGAGCGCGCATCGCCAGGTCAGCTGAGTGTGAGTGTGAT-CTGCA 1672
Db 1648 uGlyAsnSerLysAlaThrSerPheGlySerLeuArgCysGlnValLysValArgSerHl 1668
QY 1673 TGGTACTGGCCAGCCTACT-----CAAGGTATCCTAAAGTGCACGCA 1714
Db 1668 sGlyGlnAspSerProAlaValTyrLeuLysLysGlyGlnGlySerThrThrLeuGl 1688
QY 1715 GGCAGAG---CCACCCTGGGGATGGACACTGGCCCTCCTCCTCGGGAGGCC----- 1766
Db 1688 nLysArgPheGluProThrGlyPheGlnAsnMetLeuSerGlyLeuTyrAsnProIleVa 1708
QY 1767 -----TCTGGGACTCCTCGCCC-----TGCATAAAAAGAGGGTG 1801
Db 1708 lPheSerAlaSerGlyAlaAsnLeuThrAspAlaHisLeuPheCysLeuAlaCysAs 1728
QY 1802 ATTTTCTACTTGTGTATGTTGCTTTCAAATTCCTTAGTAGTACCTCCATTCAAGT 1861
Db 1728 pArgAspLeuCysCysAspGlyPheValLeuThrGlnValGlnGlyAlaIleIleCy 1748
QY 1862 T---ATTATGAGCCAGCTCAAGTTAGAGAGTAGGCTCTTCTCAGGTGGACTCTGCC 1918
Db 1748 sGlyLeuLeuSerSerProSerValLeu---LeuCysAsnValLysAspTrpMetAspPr 1767
QY 1919 AAATCATACATAAGTCAGTCAGCGGTTTTTCCAGGCCAGCGCTGTGACACAGAG 1978
Db 1767 oSerGlu-----AlaTrpAlaAsnAlaThrCysProGlyValThrTyrAspGlnGl 1784
QY 1979 ATATGGAGGGGGTGGGTTAGAGTGGGT 2009
Db 1784 u----SerHisGlnValIleLeuArgLeuGly 1793

Search completed: April 21, 2003, 11:43:07
Job time : 118.924 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:19:57 ; Search time 30.0475 Seconds

(without alignments)
6186.760 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 tctctcgtgggtggcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cn2_1/USPOO/US09743237/runat_21042003_111943_388/app_query.fasta_1.4494
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 -CGEN_1_1_93 @runat_21042003_111943_388 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1599	39.9	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
2	1251	31.2	299	1 MTL5_HUMAN	Q9v4i5 homo sapien
3	144	3.6	1183	1 DRPL_RAT	P54258 rattus norv
4	125.5	3.1	892	1 HIC1_MOUSE	Q9riy5 mus musculus
5	119.5	3.0	1185	1 DRPL_HUMAN	P54259 homo sapien
6	117.5	2.9	309	1 HXA4_CHICK	P17277 gallus gall
7	117.5	2.9	971	1 RECK_HUMAN	O95980 homo sapien
8	117.5	2.9	1700	1 BAR3_CHITE	Q03376 chironomus
9	115.5	2.9	1509	1 GSR1_HUMAN	Q9nzm4 homo sapien
10	113	2.8	643	1 CD93_RAT	Q9wt61 rattus norv
11	112.5	2.8	323	1 AMFR_HUMAN	P26442 homo sapien
12	112.5	2.8	2768	1 THYG_HUMAN	P01266 homo sapien
13	112	2.8	652	1 CD93_HUMAN	Q9npy3 homo sapien
14	111.5	2.8	946	1 IP3L_HUMAN	P27987 homo sapien
15	110.5	2.8	493	1 OC90_HUMAN	Q02509 homo sapien
16	109.5	2.7	478	1 EGR4_HUMAN	Q00911 rattus norv
17	108.5	2.7	610	1 LEM2_HUMAN	P16581 homo sapien
18	107.5	2.7	1093	1 SM5B_MOUSE	Q60519 mus musculus

c 19	107	2.6	1142	1	MGC1_HUMAN	O60732 homo sapien
c 20	106.5	2.6	1436	1	WC11_BOVIN	P30205 bos taurus
c 21	106.5	2.7	2805	1	MAPA_HUMAN	P78559 homo sapien
c 22	106	2.6	470	1	EGR2_MOUSE	P08152 mus musculus
c 23	105.5	2.6	611	1	LEM2_CANFA	P33730 canis famil
c 24	104.5	2.6	545	1	AIRE_HUMAN	O43918 homo sapien
c 25	104.5	2.6	760	1	OCT1_XENLA	P16143 xenopus lae
c 26	104.5	2.6	3149	1	TEGU_EBV	P03186 epstein-bar
c 27	104	2.6	1093	1	AF17_HUMAN	P55198 homo sapien
c 28	104	2.6	1096	1	KPCL1_ASPNG	Q00078 aspergillus
c 29	104	2.6	1680	1	FUR2_DROME	P30432 drosophila
c 30	104	2.6	1724	1	POV_HUMAN	Q9ukk3 homo sapien
c 31	104	2.6	1964	1	NTC4_MOUSE	P13695 mus musculus
c 32	104	2.6	5376	1	ZAN_MOUSE	O88799 mus musculus
c 33	103	2.6	579	1	SOC7_MOUSE	Q8vnd2 mus musculus
c 34	103	2.6	1139	1	KPCL1_TRIRE	Q99014 trichoderma
c 35	103	2.5	1801	1	LMB2_RAT	P15800 rattus norv
c 36	102	2.5	470	1	EGR2_RAT	P51774 rattus norv
c 37	102	2.5	867	1	SSPO_BOVIN	P98167 bos taurus
c 38	102	2.5	2768	1	THYG_RAT	P06882 rattus norv
c 39	102	2.5	5376	1	ZAN_MOUSE	O88799 mus musculus
c 40	101.5	2.5	3828	1	TRX_DROVI	Q24742 drosophila
c 41	101.5	2.5	4660	1	LRP2_RAT	P98158 rattus norv
c 42	101	2.5	503	1	IRF7_HUMAN	Q92985 homo sapien
c 43	101	2.5	1509	1	GSR1_HUMAN	Q9nzm4 homo sapien
c 44	101	2.5	1581	1	LMG3_MOUSE	Q9t0b6 mus musculus
c 45	101	2.5	1816	1	LMN4_HUMAN	Q16363 homo sapien

ALIGNMENTS

RESULT 1
MTL5_MOUSE
ID MTL5_MOUSE STANDARD; PRT; 295 AA.
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).
GN MTL5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77383; AAD24667.1;
DR EMBL; U67176; AAD24666.1;
DR MGD; MGI:1340029; Mtl5.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 1.
KW Spermatogenesis.

FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Alignment Scores:
Pred. No.: 9 86e-117 Length: 295
Score: 1599.00 Matches: 295
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.90% Indels: 0
DB: 1 Gaps: 0

US-09-743-237-1 (1-2241) x MTL5_MOUSE (1-295)

QY 651 ATGGTATTTTCAGCTCAAGAGGAGGCGCCAGATGCTCTGCATAGACAAGTGTGGCGG 710
DB 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 711 AGGAGCTCAAGCGCTCCATCTCTCTCCTCAGTACGATGACGAGAGTTCCTCCCTAG 770
DB 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerSerPheProGln 40
QY 771 TCAGAGCTCCCTAAGCCAAATCACAACTTACTGTGGGAAGACTTCTGCCAGTACACGCGAAG 830
DB 41 SerGluLeuProLysProMetThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 831 TTAATCTCATCACAGAGTTGATAATGGAGCTCTCCCATCAGCTGCTCAATGGGCGTCC 890
DB 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 891 TTTCCCTCTGAGCTCTCTCCAGGCGCCACCAAAATCACTCTGTCTGGGTACTGTGAC 950
DB 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGCGGGACTTCTGCAACAGCTGCGAGCTGCAACACCTGCGCGCATGAGCTC 1010
DB 101 CysPheSerSerGlyAspPheCysAsnSerCysSerCysAsnAsnLeuArgHisGluLeu 120
QY 1011 GAGCGCTTCAAGCCATAAAGCGCTGCTTGTATAGAATCCTGAAGCTTTCCCAACCAAAA 1070
DB 121 GluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAlaPheGlnProLys 140
QY 1071 ATGGGGAAGCGCTCTGGGAGCTGCTAAACTTCGACACAGCAAGGTTGCACTGTAAAG 1130
DB 141 MetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGlyCysAsnCysLys 160
QY 1131 CGCTCAGCTGCCCTGAAGAACTACTGTGAGTGTCTATGAGGCGCAAAATCACTGTCTTCC 1190
DB 161 ArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIleMetCysSerSer 180
QY 1191 ATTTGCAATGCTGTTGCAAAACTATGAGAAAGTCCAGAACGAAATAATGCTGATG 1250
DB 181 IleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluArgLysMetLeuMet 200
QY 1251 AGCACCCCACTACATGGAGCGCTTTCAGAGCAGCCATTATTGCTCCCGCAGCC 1310
DB 201 SerThrProHisTyrMetGluProGlyAspPheGluSerSerHisTyrLeuSerProAla 220
QY 1311 AAGTTCTCAGGACCTCCAAAACACTGAGAAAATAAGGAGCGCTTCTCTCTGATCTCTCGG 1370
DB 221 LysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSerCysIleSerTip 240
QY 1371 GAAGTAGTGAGGCCACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
DB 241 GluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGluAlaGluGlnGlu 260
QY 1431 CACTGTTCCCAAGCTGGCTGACACATGATCTCTGAGGAGTTTGAAGTGGCTGCTGCTGCTG 1490
DB 261 HisCysSerProSerLeuAlaGluGlnMetIleLeuGluGluPheGlyArgCysLeuSer 280
QY 1491 CAGATTCTCCACATCGAGTTCAAGTCCCAAGGGGCTGAAAATTGAG 1535
DB 281 GlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295

RESULT 2

MTL5_HUMAN
ID MTL5_HUMAN STANDARD; PRT; 299 AA.
AC Q9Y4I5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL: U86074; AAD24668.1;
CC Genes: HGNC:7446; MTL5.
CC MIM: 604374;
CC InterPro: IPR005172; CXC.
CC Pfam: PF03638; CXC; 1.
CC Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH.
SQ SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;

Alignment Scores:

Pred. No.: 1 2e-89 Length: 299
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 31.21% Indels: 4
DB: 1 Gaps: 1

US-09-743-237-1 (1-2241) x MTL5_HUMAN (1-299)

QY 651 ATGGTATTTTCAGCTCAAGAGGAGGCGCCAGATGCTCTGCATAGACAAGTGTGGCGG 710
DB 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
QY 711 AGGAGCTCAAGCGCTCCATCTCTCTCCTCAGTACGATGACGAGAGTTCCTCCCTAG 770
DB 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnTyrLeuGln 40
QY 771 TCAGAGCTCCCTAAGCCAAATCACAACTTACTGTGGGAAGACTTCTGCCAGTACACGCGAAG 830
DB 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
QY 831 TTAATCTCATCACAGAGTTGATAATGGAGCTCTCCCATCAGCTGCTCAATGGGCGTCC 890
DB 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
QY 891 TTTCCCTCTGAGCTGCTGTGCAAGGCGCCACCAAAATACTCTGTCTGGGTACTGTGAC 950
DB 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
QY 951 TGCTTCTCCAGCGGGAGCTTCTGCAACAGCTGCGAGC-----TGCAACACCTG 998


```

101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnAsnCysAsnAsnCysAsnLeu 120
102 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnAsnCysAsnAsnCysAsnLeu 120
999 CGCCATGAGCTCGAGCGCTTCAAGCCATAAAGCGGTGCTTGTAGTAAGTCTGAAGCT 1058
121 HsHsAspGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 140
1059 TTCCAAACCAAAATGGGAAAGCGCTCTGGAGCTGCTAAACTTCGACACAGCAAGG 1118
141 PheGlnProLysIleGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 160
1119 TGAAGTGTAAAGCGCTAGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1178
161 CysAsnCysArgArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGln 180
1179 ATGTGTTCTTCCATTTGCAATGCTGCTGCAAAAGCTATGAGAAAGTCCAGAACGA 1238
181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGlu 200
1239 AAAATGCTGATGAGCACACCCACTACATGGAGCGCTGGGACTTTGAGAGCAGCC 1298
201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHis 220
1299 TTGTCCTCCCAAGCAAGTCTCAGAGCTCCAAAGCTGAGAAAGTAGGAGCGCTTCT 1358
221 LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgProSerSer 240
1359 TGTATCTCTGGAAGTGTAGGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1418
241 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGlu 260
1419 GCAGAGCAGGAGCAGTGTGCTCCCAAGCTGCTGAGCAGATGCTGAGGAGTTTGA 1478
261 AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluPheGly 280
1479 AGTGCTCTGCGAGATTTCCACATCGAGTTCAGTTCAGGAGGCTGAAATTTGAG 1535
281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 3
DRPL_RAT
ID DRPL_RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RA "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT homologue."
RT Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidoluysian atrophy (DRPLA) in rat."
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -!- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -!- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31777; AAA80337.1; -
DR EMBL; X89453; CAA61623.1; -
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT CONFLICT 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 P -> L (IN REF. 2).
FT CONFLICT 689 689 F -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCADF9B1F CRC64;

Alignment Scores:
Pred. No.: 0.00232 Length: 1183
Score: 144.00 Matches: 140
Percent Similarity: 27.98% Conservative: 48
Best Local Similarity: 20.83% Mismatches: 177
Query Match: 3.56% Indels: 308
DB: 1 Gaps: 28

US-09-743-237-1 (1-2241) x DRPL_RAT (1-1183)
QY 1997 CCCGACCCCTCCCATATCTCTGTACAGCGCTGGCTGGAAAAACCCCTGATGGCA 1938
DB 171 ProaspSerIleProArgGlnProGluSerGlyPheGlu---ProHisProSerValPro 189
QY 1937 CTTGACTTGTATGTGATTTGGCAGAGTCCACCTGAAGAAGAGCCTCTCTAACTTG 1878
DB 190 ProThrGlyTyr----- 193
QY 1877 AGCTGGCTCATANTAACTGAATGGAGTACTACTACTAAGCAATTTGAAAGCAACACATA 1818
DB 194 -----HisAlaProMetGluProProThrSerArgLeuPheGlnGlyProProPro 210
QY 1817 ACAACAAGTAGAAAATCACCTCTTTTAT-----GCAGGGCAGGAGTCCCCAGA 1767
DB 211 GlyAlaProProHisProGlnLeuTyrProGlySerAlaGlyGlyGlyValLeuSer 230
QY 1766 GGGCTCTCCCGAGGACAGGAGGCGCAGTGT-----CCATCCCGCAGG 1725
DB 231 GlyProProMetGlyProLysGlyGlyAlaAlaAlaSerValGlyPro-ProSerG 250
QY 1724 TGGCTCTGCTGCTTGCCTTGTAGGATACCTTGTAGTGGCTGGCGGAGTACCAGTGCAGATC 1665
DB 250 yGlyLys-----GlnHisProProProThrThrProIleProIleSerSe 265
QY 1664 A----- 1664
DB 265 rSerGlyAlaSerGlyAlaProProAlaLysProProAsnThrProValGlyAlaGlyAs 285
QY 1664 ----- 1664
DB 285 nLeuProSerAlaProProProAlaThrPheProHisValThrProAsnLeuProProPr 305

```

QY 1663 -----CTCAGACCT-----CACAGCTGACCTGCCATGATGGCGGCC 1626
Db 305 oProAlaLeuArgProLeuAsnAlaSerAlaSerProGlyMetGlyAlaGlnPr 325
QY 1625 TCTG-----CTCCTCC 1614
Db 325 oileProGlyHisLeuProSerProHisAlaMetGlyGlnGlyMetSerGlyLeuProPr 345
QY 1613 TGGG-----CACAGCTTCTCGGTCCAGATT 1587
Db 345 oGlyProGlyLysGlyProThrLeuAlaProSerProHisProLeuPro-ProAlaSerS 365
QY 1586 CCCAGGCTGAGGCTGCCAGCATTCCTTTTACCAGCTTGCACGCTACTCAATTTT 1527
Db 365 erSerAlaProGlyProProMetArgTyrProTyrSerSerSerSerSerVala 385
QY 1526 CAGCCCTTGGACTTGAACCTGATGGAGAACTGCGACAGGACCTTCCAACTCCT- 1468
Db 385 laAla-----SerSerSerSerAlaAlaThrSerGlnTyrProA 399
QY 1468 -----CTGCTCTGCTTCT- 1468
Db 399 laSerGlnThrLeuProSerTyrProHisSerPheProProProThrSerMetSerValS 419
QY 1467 -----CCAGGATCATCTGCTCAGCCAACTTGGGGAACAGTGCCT- 1429
Db 419 erAsnGlnProProLysTyrThrGlnProSerLeuProSerGlnAlaValTyrSerGlnG 439
QY 1428 -----CTGCTCTGCTTCT- 1407
Db 439 lyProProProProProProProTyrGlyArgLeuLeuProAsnAsnThrHisProG 459
QY 1406 GGCA-----GCAGGAGGACATGTGGCTCCACTACTTCCAGGAGATACA 1359
Db 459 lyProPheProProThrGlyGlyGlnSerThrAlaHisProProAlaProAlaHisHis 479
QY 1358 GGAGAGGCTGCCTATTTTCTCAGTTTGGAGGCTCTGAGAACTTGGCTGGGACAA 1299
Db 479 lsHisGlnGln----- 482
QY 1298 ATATGCTGCTCTCAAGTCCCG- 1273
Db 483 -----GlnGlnGlnProGlnProGlnProGlnProGlnHisHisGlyA 499
QY 1272 -----GCTCCATGATGAGTGGGTGCTCATCAGCATTTTCTGCTGACTTTCTTC 1221
Db 499 snSerGlyProProProProGlyAla----- 507
QY 1220 ATAGTTTTTGAACCAATGATTTGCAATGGAGAACACATGATTTTGGCTCATAGCA 1161
Db 508 -----TyrProHisProL 512
QY 1160 CTCACAGTACTTCTCA- 1137
Db 512 euGluSerSerAsnSerHisHisAlaHisProTyrAsnMetSerProSerLeuGlySerL 532
QY 1136 TGAGCGCTTACAGTTG-----CACCTTTGCTGTGTCGAA 1102
Db 532 euArgProTyrProProGlyProAlaHisLeuProProSerHisGlyGlnValSerTyrS 552
QY 1101 GTTTAGAGCTCCAGAGGCTTTCCCATTTTGGTTGGAAGCTTCAAGATTCTTAT 1042
Db 552 erGlnAlaGlyProAsnGlyProProValSerSerSerSerAsnSerSerGlySerS 572
QY 1041 CAACACAGCTTTATGGCTTTGAGGCTGCGAGCTCATGCCAGCTTGTTCAGCTGC 982
Db 572 erGlnAlaAlaTyrSerCysSerHisProSerSer----- 584
QY 981 AGCTGTTGCAGAGTCCCGCTGGAGAGCAGTACAGTACCCACAGACAGATTATTTGG 922
Db 584 ----- 584

QY 921 GTGGCCCTTCAGAGCAGGTCCTCAGAGGAAAGCAGGCCCATTTGACAGCTGATGGGAGAG 862
Db 585 -----GlnGlyProGlnGlyAlaSerTyrProPhe-----P 595
QY 861 CTCATTATCAACTGTGTGATGAGATTAACTTCGCTGCTACTGGCAGAGTCTTCCCA 802
Db 595 roProVal-ProProfile-----ThrThrSerSerAlaThrLeuSerThrValile--- 611
QY 801 CTAAGTGTGCTTGGCTTAGGAGCTCCTCAGTTCAGGGAACCTGCTGCTGCTCATCTPACT 742
Db 612 -----AlaThrValaSerSerProAlaGlyTyrLysThr 623
QY 741 GAGGAAGCAGATGAGCGCTTTGAGCTCCTCGCGCCACAGTTGCTTCATGACAGCATCT 682
Db 624 AlaserProGly-----ProProGlnTyrSerLysArgAlaProSerPro 639
QY 681 GGGCGCTCTTCAGCTGACAAATCACCATGGGCTGGAGTCTTTCTTCGAGGCGCAGC 622
Db 640 GlyserTyrLysThrAla-----ThrProGlyTyrLysProGlySerProProSer 657
QY 621 TGAGGCTCTCTGCTGCTGATGGAACTTGCACAGGATTCCTGAGCCAGAG 562
Db 658 PheArgThrGlyThrProProGlyTyrArgGlyThr-----SerProProAla 673
QY 561 GGGCGTGAAGTGGCTCTTCAGGCGTCCG-----CTGCGACACTACCACTG 511
Db 674 GlyProGlyThrPheLys--ProGlySerProThrValGlyProGlyProLeuProProA 693
QY 510 CT-----TCTTTGATTTCAACACTT 491
Db 693 laglyProSerSerLeuSerSerLeu 701
RESULT 4
HICL_MOUSE STANDARD; PRT; 892 AA.
ID HICL_MOUSE Q9RIY5; Q9RIY6; Q9R280;
AC Q9RIY5; Q9RIY6; Q9R280;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypermethylated in cancer 1 protein (Hic-1).
GN HIC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutharia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC STRAIN=129/Sv, and Swiss Webster; TISSUE=Embryo;
RX MEDLINE=99172081; PubMed=10072440;
RA Grimm C., Spoerle R., Schmid T.E., Adler I.-D., Adamski J.,
Schughart K., Graw J.;
RT "Isolation and embryonic expression of the novel mouse gene Hic1, the
homologue of Hic1, a candidate gene for the Miller-Dieker syndrome.";
RL Hum. Mol. Genet. 8:697-710(1999).
RN [2]
RP SEQUENCE FROM N.A., DISEASE, AND TISSUE SPECIFICITY.
RX MEDLINE=20122251; PubMed=10655551;
RA Carter M.G., Johns M.A., Zeng X., Zhou L., Zink M.C., Mankowski J.L.,
Donovan D.M., Baylin S.B.;
RT "Mice deficient in the candidate tumor suppressor gene Hic1 exhibit
developmental defects of structures affected in the Miller-Dieker
syndrome.";
RL Hum. Mol. Genet. 9:413-419(2000).
RN [3]
RP SEQUENCE OF 179-338 FROM N.A.
RC STRAIN=129/Sv; TISSUE=Liver;
RX MEDLINE=99297610; PubMed=10371200;
RA Guerardel C., Delour S., Leprince D.;
RT "Evolutionary divergence in the broad complex, trantrack and bric a
brac/poxviruses and zinc finger domain from the candidate tumor
suppressor gene hypermethylated in cancer.";
RL FEBS Lett. 451:253-256(1999).
CC -!- FUNCTION: Putative transcription factor. May act as a tumor

CC suppressor. May be involved in development of head, face, limbs
 CC and ventral body wall.
 CC -!- SUBUNIT: Interacts with CtBP (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- ALTERNATIVE PRODUCTS: A number of isoforms may be produced by
 CC alternative splicing.
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed with highest levels in
 CC heart and lung.
 CC -!- DEVELOPMENTAL STAGE: Expression is first detected in the embryo
 CC after 9 days post coitum. In the embryo, expression is found in
 CC restricted regions of somite derivatives, limb anlagen and cranio-
 CC facial mesenchyme. In the fetus, it is additionally expressed in
 CC mesenchymes apposed to precartilaginous condensations, at many
 CC interfaces to budding epithelia of inner organs, and weakly in
 CC muscles.
 CC -!- DISEASE: Defects in HIC1 are the cause of perinatal death with
 CC serious developmental anomalies, including acrania, exencephaly,
 CC cleft palate, omphalocele, craniofacial and limb anomalies.
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.
 CC -!- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -!- SIMILARITY: CONTAINS 5 C2H2-TYPE ZINC FINGERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF036334; AAD30654.1; -;
 CC EMBL: AF036582; AAD30655.1; -;
 CC EMBL: AJ332691; CAB44493.1; -;
 CC HSSP: P08046; 1A1H.
 CC MGD: MGI:1338010; Hic1.
 CC InterPro: IPR000210; BTB_POZ.
 CC InterPro: IPR000822; Znf_C2H2.
 CC Pfam: PF00651; BTB; 1
 CC Pfam: PF00096; zf-C2H2; 5.
 CC ProDom: PD000003; Znf-C2H2; 1.
 CC SMART: SM00225; BTB; 1.
 CC SMART: SM00355; Znf-C2H2; 5.
 CC PROSITE: PS50097; BTB; 1.
 CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
 CC PROSITE: PS5015; ZINC_FINGER_C2H2_2; 5.
 CC Multigene family; Nuclear protein; Transcription regulation;
 CC DNA-binding; Zinc-finger; Metal-binding; Repeat;
 CC Developmental protein; Alternative splicing.
 CC FT DOMAIN 1 172 ARG/PRO/GLY-RICH.
 CC FT DOMAIN 206 269 BTB.
 CC FT DOMAIN 271 278 POLY-ALA.
 CC FT DOMAIN 319 326 POLY-GLY.
 CC FT DOMAIN 354 358 POLY-PRO.
 CC FT DOMAIN 596 777 ZINC FINGERS.
 CC FT ZN_FING 596 623 C2H2-TYPE.
 CC FT ZN_FING 666 693 C2H2-TYPE.
 CC FT ZN_FING 694 721 C2H2-TYPE.
 CC FT ZN_FING 722 749 C2H2-TYPE.
 CC FT ZN_FING 750 777 C2H2-TYPE.
 CC FT SITE 400 404 BINDING TO CTBP.
 CC FT CONFLICT 242 242 N -> S (IN REF. 1: AAD30654).
 CC SQ SEQUENCE 892 AA; 94320 MW; BBD16AA051995740 CRC64;
 CC
 CC Alignment Scores:
 CC Pred. No.: 0.0602 Length: 892
 CC Score: 125.50 Matches: 173
 CC Percent Similarity: 31.49% Conservative: 49
 CC Best Local Similarity: 24.54% Mismatches: 233
 CC Query Match: 3.10% Indels: 251
 CC Db: 1 Gaps: 41

US-09-743-237-1 (1-2241) x HIC1_MOUSE (1-892)

Qy 1817 ACAACAGTACAAATACACCTCTTTTATCGAGGCGAGGAGTCCCGAGAGGCGCTCCC 1758
 Db :||| :|||:||||| ||| :|||:||||| |||||
 Db 38 SerThrGlnHisLysThrProArgPhe-----MetGlyLeuPro---GlyProIle 53
 Qy 1757 CAGGACAGGAGGCGCAGTGTCCATCCCGAGG-----TGCGCTGCTGCTTGCAT 1707
 Db :|||:||||| ||| :|||:||||| |||
 Db 54 Ser-----GluSerProGlyGlyHisIleTrpLysArgAla----- 65
 Qy 1706 CTTTAGTACCTTGTAGTAGCTGGCCAGTACCAGTACCTCAGACCTCAGAGCTG 1647
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 66 -----AsnLeuGlnAspLeuLeuSerSerAlaSerAlaSerLeuLeuAlaGlnVal 82
 Qy 1646 ACCTGGCCATGATCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1587
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 83 -----CysAlaArgGlyArgSerProAlaAlaHisSer----- 93
 Qy 1586 CCCAGGCTGAGGCTGGCCAGAGCATTTCCCTTTTACCAGCTTGCACGCTACTCAATTTT 1527
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 94 ProArgValAlaAlaArgTrpArgHisGlyArgGlySerValCys-Arg-----Ph 110
 Qy 1526 CAGCCCTTGGACTTGAACTCGATGTGGAGAACTTGGCAGAGGACCTTCCAACTCCTC 1467
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 110 eGlyProLeuGlnIleArgVal-CysGlyLysArgGly----- 122
 Qy 1466 CAGCATCATCTGCTCAGCCCAAGCTTGGGACAGTGTCTCTCTCTCTCTCTCTCTCTCT 1407
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 123 --GlyAlaGluThrArgProGlyArgGlyGluAspGlyProAlaArgGlnThrAspArg 142
 Qy 1406 GCCCAGGAGGAGG-----CACATGTGGCTCCACTACTTCCAGGA 1365
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 142 LysProGlyGlyArgAlaAlaHisCysSerHisVal---ProProTrpIle 158
 Qy 1364 GATACAGGAGAGGCTGCT 1305
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 159 -----ArgArgGlnPro----- 162
 Qy 1304 GGACAATAATGGCTGCTCTCAAGTCCCGAGGCTAGTGGGGGTGCTCATCAG 1245
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 163 -----GlyProGlyLeuProThrCysProProGlyGluCysAlaGlyGln----- 177
 Qy 1244 CATTTTCTGTTCTGACTTCTCT-----CATAGT----- 1216
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 178 -----MetMetLeuAspThrMetGluAlaProGlyHisSerArgGlnLeuLeuGlnL 196
 Qy 1215 -----TTTGCAGCAATGATTTTGCAGCAATGATTTTGCAGCAATGATTTTGCAG 1179
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 196 euAsnAsnGlnArgThrLysGlyPheLeuCysAspValIleIleValValGlnAsnAlaL 216
 Qy 1178 GATTTTGCCTCATAGCACTCAGTACTTCTTTCAGGAGGCTGAGCGCTTACAGTTGCA 1119
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 216 euPheArgAlaHisLysAsnValLeuAlaAlaSer-----SerAlaTyrLeuLys 233
 Qy 1118 CCCTTTGCTGTGTGCAAGTTTATAGCAGCTCCAGAGCGGCTT----- 1078
 Db :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 233 erLeuVal-----ValHisAspAsnLeuLeuAsnLeuAspHisAspM 247
 Qy 1077 -----TCCCATTTTGTGGTGGAGCTTCAAGTATTTCTATCAGACACAGCGCTTATGGC 1023
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 247 etValSerProAlaValPheArgLeuValLeuAspPhe-IleTyrThrGly----- 263
 Qy 1022 TTTGAAGGCTCGAGCTCATGGCGAGGTTGTTTCAGCTGCGAGCTGTTGCGAGAGTCCCC 963
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 264 -----ArgLeuThrAspSerValGluAlaAlaAlaAlaAlaValAlaAla 279
 Qy 962 GCTGGAGAGCAGTCA-----CAGTACCCAGA 936
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 280 GlyAlaGlu-ProSerLeuGlyAlaValLeuAlaAlaAlaSerTyrLeuGlnIleProAs 299
 Qy 935 CAGAGTTATTTTG-----GGTGGCCCTTGC-----AGAGC 906
 Db :||| :|||:||||| ||| :|||:||||| ||| :|||:||||| ||| :|||:||||| |||
 Db 299 pLeuValAlaLeuCysLysLysArgLeuLysArgHisGlyLysTyrCysHisLeuArgGln 319

AC P17277;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-A4 (Chox-1.4).
 GN HOXA4 OR CHOX-1.4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90245562; PubMed=1970866;
 RA Sasaki H., Yokoyama E., Kuroiwa A.;
 RT "Specific DNA binding of the two chicken Deformed family homeodomain
 proteins, Chox-1.4 and Chox-a";
 RL Nucleic Acids Res. 18:1739-1747(1990).
 RN [2]
 RP SEQUENCE OF 207-273 FROM N.A.
 RC STRAIN=Comet Hubbard hybrid;
 RX MEDLINE=90326535; PubMed=1973835;
 RA Scotting P.J., Hewitt M., Keynes R.J.;
 RT "Isolation and analysis of chick homeobox cDNA clones";
 RL Nucleic Acids Res. 18:3999-3999(1990).
 CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 SPECIFIC POSITIVE IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -!- FUNCTION: BINDS TO SITES IN THE 5'-FLANKING SEQUENCE OF ITS CODING
 REGION WITH VARIOUS AFFINITIES. THE CONSENSUS SEQUENCES OF THE
 HIGH AND LOW AFFINITY BINDING SITES ARE TAATGA(C/G) AND CTAATTT.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- MISCELLANEOUS: THE PROLINE STRETCH WORKS AS A PART OF THE
 TRANSCRIPTIONAL ACTIVATION DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE ANTP HOMEBOX FAMILY.
 CC "DEFORMED" SUBFAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X52670; CAA36896.1; -
 DR EMBL: X52747; CAB57949.1; -
 DR PIR: S09257; S09257.
 DR PIR: S10883; S10883.
 DR HSP: P02833; 9ANT.
 DR TRANSPAC: T00128; -
 DR InterPro: IPR001827; Antennapedia.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR PRINTS: PR00025; ANTENNAPEIDIA.
 DR PRINTS: PR00024; HOMEBOX.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 DR PROSITE: PS00032; ANTENNAPEIDIA; 1.
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 Transcription regulation.
 FT DOMAIN 38 148 PRO-RICH.
 FT DOMAIN 188 193 ANTP-TYPE HEXAPEPTIDE.
 FT DNA_BIND 209 268 HOMEBOX.
 SQ SEQUENCE 309 AA; 33478 MW; 8898A41502319341 CRC64;
 Alignment Scores:
 Pred. No.: 0.202 Length: 309
 Score: 117.50 Matches: 58
 Percent Similarity: 36.36% Conservatives: 14
 Best Local Similarity: 29.29% Mismatches: 81

Query Match: 2.90% Indels: 45
 DB: 1 Gaps: 9
 US-09-743-237-1 (1-2241) x HXAA4_CHICK (1-309)
 QY 1829 AGCAACACATACACAAAGTAGAAATCACCCTCTTTTATGAGGCGAGGAGTCCCC 1770
 Db 24 ThrGlnHisSerGlySerAlaGlySerSerAlaSerTyrHisProHisHisProHisPro 43
 QY 1769 AGAGGGCTCCCGCAGGACAGGAGGCGCAGTGTCATCCGCCAGGAGTCTGCTGCTGCT 1710
 Db 44 HisAlaProProPro-----Pro-ProProProProProHisLe 57
 QY 1709 GCATTTAGGATACCTTGTAGTGGCCAGTACCATGAGATCAGATCAGACCTCAGAC 1650
 Db 57 uHisAlaHisPro-----GlyProAlaLeuProGluTyrPheProArgProArg 75
 QY 1649 CTGACCTGGCCATGATGCGGGGCTCTGCTCCCTCCGCGGACAGGAGTCTGCTGCTG 1590
 Db 75 gGluProGlyTyrGlnAla---ProAlaAlaProProGlyProProGlyProPro----- 92
 QY 1589 ATTCCCGAGGCTGAGGCTTGCACAGGAGCATTCCTCTTTTACAGCTTGCAGCTACTCAAT 1530
 Db 93 ----ProGluAlaLeuTyrProAlaGlnAlaProSerTyrProGlnAlaProTyrSer-T 111
 QY 1529 TTTTCAGCCCTTGGACTTGAACCTCGATGTGGAGAATCTGCGACAGGACCTTCCAAATC 1470
 Db 111 yrSerSer-----AlaGlySerAlaAlaProGlyProGluGlnProp 125
 QY 1469 CTCAGGATCATCTGCTCAGCCAAAGCTTGGGAAAGCTGCTCCTGCTGCTCTCTCTCACC 1410
 Db 125 roProGlyAlaSerProProPro-----ProProAlaAlaGlyHisP 140
 QY 1409 CTGGGCGAGGACGAGGACACATGTGGCTCCACTACTTCCC-----AGGA 1365
 Db 140 roGlyProAla-----GlnProLeuLeuProGlyHisAlaLeuGlnArg 155
 QY 1364 GATACAGGAGGAGGCTGCTCTATTTTTCAGTGTGGAGGCTCCTGAGAACTTGGCTGG 1305
 Db 155 rgCysGluAlaAlaProAla-----AlaGlyAlaG 165
 QY 1304 GGACAAATATGCTGCTCTCAAGTCCCGAGGCTCCATGATGAGTGGGGTG 1255
 Db 165 lyThrGlyProGlyCysAlaLeuLeuProAspTyrSerLeuProGlyLeu 181
 RESULT 7
 RECK_HUMAN
 ID RECK_HUMAN STANDARD; PRT; 971 AA.
 AC O95980; Q8WX37;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Reversion-inducing cysteine-rich protein with Kazal motifs precursor
 DE (hRECK) (Suppressor of tumorigenicity 15) (STI5).
 GN RECK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND N-GLYCOSYLATION.
 RC TISSUE=Fibroblast;
 RX MEDLINE=99007295; PubMed=9789069;
 RA Takahashi C., Sheng Z., Horan T.P., Kitayama H., Maki M., Hitomi K.,
 RA Kitaura Y., Takai S., Sasahara R.M., Horimoto A., Ikawa Y.,
 RA Ratzkin B.J., Arakawa T., Noda M.;
 RT "Regulation of matrix metalloproteinase-9 and inhibition of tumor
 invasion by the membrane-anchored glycoprotein RECK";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13221-13226(1998).
 RN [2]
 RP SEQUENCE OF 363-971 FROM N.A.
 RA Kimberley A.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Negatively regulates matrix metalloproteinase-9 (MMP-9)
 CC by suppressing MMP-9 secretion and by direct inhibition of its
 CC enzymatic activity. RECK down-regulation by oncogenic signals may
 CC facilitate tumor invasion and metastasis. Appears to also
 CC regulate MMP-2 and MT1-MMP, which are involved in cancer
 CC progression.
 CC -1- SUBUNIT: Interacts with MMP-9.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: Expressed in various tissues and untransformed
 CC cells. It is undetectable in tumor-derived cell lines and
 CC oncogenically transformed cells.
 CC -1- SIMILARITY: CONTAINS 3 KAZAL-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: D50406; BAA34060.1; -
 CC EMBL: AL158830; CAD13384.1; -
 CC Genew: HGNC:11345; RECK.
 CC MIM: 605227; -
 CC HSP; P80424; IAN1.
 CC InterPro: IPR002350; kazal.
 CC SMART: SM00280; KAZAL; 3.
 CC SMART: SM00011; VWC_def; 1.
 CC PROSITE: PS00282; KAZAL; 1.
 CC Signal: Glycoprotein; GPI-anchor; Serine protease inhibitor;
 KW Membrane; Anti-oncogene; Repeat.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 942 REVERSION-INDUCING CYSTEINE-RICH PROTEIN
 FT WITH KAZAL MOTIFS.

FT PROPEP 943 971 REMOVED IN NATURE FORM (POTENTIAL).
 FT LIPID 942 942 GPI-ANCHOR (POTENTIAL).
 FT DOMAIN 632 677 KAZAL-LIKE 1.
 FT DOMAIN 708 750 KAZAL-LIKE 2 (DEGENERATE).
 FT DOMAIN 753 787 KAZAL-LIKE 3 (DEGENERATE).
 FT DOMAIN 37 338 5 X KNOT REPEATS.
 FT REPEAT 37 84 KNOT 1.
 FT REPEAT 104 141 KNOT 2.
 FT REPEAT 151 197 KNOT 3.
 FT REPEAT 216 263 KNOT 4.
 FT REPEAT 292 338 KNOT 5.
 FT DISULFID 635 654 BY SIMILARITY.
 FT DISULFID 633 658 BY SIMILARITY.
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 971 AA; 106456 MW; 173D47D6AE6F834 CRC64;

Alignment Scores:
 Pred. No.: 0-257 Length: 971
 Score: 117-50 Matches: 84
 Percent Similarity: 30.84% Conservative: 52
 Best Local Similarity: 19.05% Mismatches: 141
 Query Match: 2.93% Indels: 164
 DB: 1 Gaps: 23

US-09-743-237-1 (1-2241) x RECK_HUMAN (1-971)

QY 660 TGTACGTGAAGAGCGCGCCAGATCCTCTGCATACACAACTGTGGCGGAGGAGCTC 719
 Db 422 CysHisSerLysSerArgGlySerIleCysLysSerAspCys----- 436
 QY 720 AAGCGCTCCATCTGCTTCTCTAGTAGGACGACGAGGAGTTCCCTCAGTCAGAGCTC 779
 Db 437 -----ValGluIleLeuLysLysCysGlyAspGlnAsnLysPheProGluAspHisThr 454

QY 780 CCTAAGCCCAATGACAACTTTAGTGGGA-----AGACTTCTGCCAGTA 821
 Db 455 AlaGluSerIleCysGluLeuLeuSerProThrAspSerLeuLysAsnCysIleProLeu 474
 QY 822 -----CCAGCGAAGTTA---AATCTCATCACACAGGTTGATAATGAGCT 863
 Db 475 AspThrTyrlleuArgProSerThrLeuGlyAsnIleValGluGluValThrHisProCys 494
 QY 864 CTCCCATCA-----GCTGTCATATGGGCTGCTTTCCTCT 899
 Db 495 AsnProAsnProCysProAlaAsnGluLeuCysGluValAsnArgLysGlyCysProSer 514
 QY 900 GGA---CCTGCTCTGAAGGCCACCAATAACTCTGCTGGTACTGACTGCTTC 956
 Db 515 GlyAspProCysLeu-----ProTyrlleuGlyValGlnGly---CysLysLeuGly 530
 QY 957 TCCAGGGGACATCTGCAACAGCTGCAGCTGCAACACCTGCCCATGAGCTCGAGCGC 1016
 Db 531 GluAlaSerAspPhe----- 535
 QY 1017 TTCAAAGCCCAATAAGGGGTCTTTGATAGAAATCTTCAACCTTCCAAACCAAAATGGGG 1076
 Db 536 -----IleValArgGlnGlyThrLeuIleGlnValProSerSer 548
 QY 1077 AAAGCGCTCTGGGAGCTGCTAAACTTCGACACAGCAAGGGTGCAACTCTAAAGCGCTCA 1136
 Db 549 AlaGlyGluValGlyCysTyrlleu-----CysSerCysGlyGlnSer 563
 QY 1137 GGCTGCTGAAG----- 1148
 Db 564 GlyLeuLeuGluAsnCysMetGluMetHisCysIleAspLeuGlnLysSerCysIleVal 583
 QY 1149 -----AACTACTGTGAGTGC 1163
 Db 584 GlyGlyLysArgLysSerHisGlyThrSerPheSerIleAspCysAsnValCysSerCys 603
 QY 1164 TATGAGCGCAAAATCATGCTGCTCTTC----- 1190
 Db 604 PheAlaGlyAsnLeuValCysSerThrArgLeuCysLeuSerGluHisSerSerGluAsp 623
 QY 1191 -----ATTGCAAAATGCATT-----GCTTGC 1211
 Db 624 AspArgArgThrPheThrGlyLeuProCysAsnCysAlaAspGlnPheValProValCys 643
 QY 1212 -----AAAACTATGAAGAAATGCCAGAACGAAATGCTGATGACACACC 1259
 Db 644 GlyGlnAsnGlyArgThrTyrlleuAlaCysIleAlaArgCysValGlyLeuGlnAsp 663
 QY 1260 CACTACATGAGCGCTGGGAGCTTTGAGAGCAGCCATTATTGTCGCCGCAAGTTCTCA 1319
 Db 664 HisGlnPheGluPheGlySerCysMetSerLys-----AspProCys----- 677
 QY 1320 GGACCTCCAAACTGAGAAAAAATAGCAGGCC----- 1352
 Db 678 AsnProAsnProCysGlnLysAsnGlnArgCysIleProLysProGlnValCysLeuThr 697
 QY 1353 -----TTCCTCTGATCTCTGGGAAGTAGTGAGGCCACATGCTGCTGCTG 1400
 Db 698 ThrPheAspLysPheGlyCysSerGlnTyrlleuGluCysValProArgGlnLeuAlaCysAsp 717
 QY 1401 CTGGCCCGAGGCTGAGGAAGCAGAGCAGGACTGTTTCCCA-----AGCTTG 1448
 Db 718 GlnValGlnAspProValCysAspThrAspHisMetGluHisAsnAsnLeuCysThrLeu 737
 QY 1449 GCTGAGCAGATGATCTCTGGAGGAGTTTGAAGGTGCTGCGAGATTTCCCATCGAG 1508
 Db 738 TyrGln-ArgGlyLysSerLeuSerTyrlleuGlyProCysGlnProPheCysArgAlaLath 757
 QY 1509 TTCA-----AGTCCAAAGGGCTGAAATTTAGTAGCTGCTGCTGCTGCTGCT 1551
 Db 757 rGluProValCysGlyHisAsnGlyLeuThrTyrlleuSerSerValCysAlaAlaTyrlleu 777

Qy 1407 CAGGGTGAAGAGCAGAGGACACTGTTCCCAAGCTTGCTGAGCAGATGATCTG 1466
 Db 478 -----GlucyGlnProGlyMetProProGlyGlyGlnThrTr 493
 Qy 1467 GAGGAGTTTGAAGGCTGCTGCGAGATTCTCCACATCGAGTTCAAGTCCAAGGCGTG 1526
 Db 493 pIleSerAspLysCysLysCysGlu-CysSerPro---ThrIleThrCysGlnAlaProG 512
 Qy 1527 AA-----AATTGAGTACGCTGCAAG-----CTGTAAGGGGA 1559
 Db 512 inIleLeuAsnThrCysGluCysLysCysProValAsnMetLeuAlaGlnLysG 532
 Qy 1560 ATGCTGTGGCAAGCTCAGCCCTGGGAATCTGCACGAGGAAGCTGCTGC----- 1610
 Db 532 LuLysCysLysSerProArgGlnThrThrAspSer---LysCysLeu-CysGluCysSer 550
 Qy 1611 -----CCAGGAGGAGCAGAGCGCGGCATCATGCGCAGGTGAG 1649
 Db 551 ThrThrProAlaThrCysGluGlyLysGlnThrTrpCysGlyGlu 565

RESULT 9

GSRL_HUMAN STANDARD; PRT; 1509 AA.
 AC Q9NZM4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glioma tumor suppressor candidate region gene 1 protein.
 GN GLTSCR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE-20175430; PubMed-10708517;
 RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
 RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
 RA Scheichauer B.W., Louis D.N., Jenkins R.B.;
 FT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
 region.";
 RT Genomics 64:44-50(2000).
 RL CC
 CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
 CC placenta, skeletal muscle, and pancreas, and at lower levels in
 CC lung, liver, and kidney.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF182077; AAF62874.1;
 CC Genew; HGNC:4332; GLTSCR1.
 CC MIN; 605690;
 DR DOMAIN 37 45 POLY-GLY.
 FT DOMAIN 884 889 POLY-PRO.
 FT DOMAIN 1214 1225 POLY-SER.
 FT DOMAIN 1282 1286 POLY-PRO.
 FT DOMAIN 1294 1304 POLY-PRO.
 SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F43CE6821 CRC64;

Alignment Scores:

Pred. No.: 0.404 Length: 1509
 Score: 115.50 Matches: 140
 Percent Similarity: 32.03% Conservative: 56
 Best Local Similarity: 22.88% Mismatches: 217
 Query Match: 199
 DB: 1 Gaps: 29

US-09-743-237-1 (1-2241) x GSRL_HUMAN (1-1509)

Qy 2065 GTTTTCTCACTTCAAGCAGAAATACAGAGAAAGAAAAACGCAAAATATCCAAACACCC 2006
 Db 783 IlePheValIleGlnAsnGlnLeuGlyValProProAlaSerAsnPro---AlaPro 801
 Qy 2005 AGCTCTTACCCGAGCCCTCCCATATCTCTGTCCACAGGCTGCTGGAAAAACCCCT 1946
 Db 802 ThrAlaProGlyProProGlnProLeuArgProGlnSerGlnPro-----Pro 818
 Qy 1945 GATGGCCACCTGACTTGTATGTGATTTGGGAGAGTCCACCTGAAGAGACCCTAGCTCT 1886
 Db 819 GluGlyProLeu-----ProProAlaProHisLeuProProSerSer 832
 Qy 1885 CTAACTTTCAGCTGCTCATATACTTGAATGAGGTACTACTAAGCAATTTGAAGCA 1826
 Db 833 ThrSer-----SerAlaValAlaSerSerSerGluThr 843
 Qy 1825 AACACATAACAAGTAGAAATACCCCTCTTTTATGCAGGCGAGGAGTCCCGCAG 1766
 Db 844 SerSer-----ArgLeuProAla 849
 Qy 1765 GGCCTCCCGAGGACAGAGGCGCAGTGTCCATCCCGAGGTGGCTCTGCTGTGTCAC 1706
 Db 850 ProThrProSerAspPheGlnLeuGlnPheProProSerGlnGly----- 864
 Qy 1705 TTTAGGATACCTTTCAGTAGGCTGCGCCAGTACCAGTCCAGATCAGATCAGACCT- 1655
 Db 865 -----ProHisLysSerProProProProProProProProProProThr 874
 Qy 1654 ---CACAGCTCACCTGGCCATGATGCGCGCTCTGCTCTCCCTCCCT---GGCACCACGTT 1601
 Db 875 LeuHisLeuValProGluProAlaAlaProProProProProProProProProPheGlnMet 894
 Qy 1600 CCTCGTGCAGATTCCCGAGGCTG-----AGG 1574
 Db 895 ValThrThrProPheProAlaLeuProGlnProLysAlaLeuLeuGluArgPheHisGln 914
 Qy 1573 CTTCGCCACAGCAATCCCTTTTACCAGTTTCAGCGTACTCAATTTTTCAGCCCTTGAC 1514
 Db 915 ValProSerGlyIleIleLeuGlnAsnLysAla-GlyGlyAlaProAlaProGlnTh 934
 Qy 1513 TTGAATCGATGTGAGAAATCTGCGACAGGACCTTCCAAACTCTCTCCAGGATCTGC 1454
 Db 934 rSerThrSerLeuGlyPro-----LeuThrSerProAlaAlaSerVa 948
 Qy 1453 T-----CAGCCAAGCTGGGAGACAGTGTCTCTGCTCTGCTCTCACCC 1409
 Db 948 lLeuValSerGlyGlnAlaProSerGlyThrProThrAlaProSerHisAlaProAlaPr 968
 Qy 1408 TGGGCCACGAGGACAGCATGTGCGCTCCACTACTTCCAGGAGATACAGGAGAGGCC 1349
 Db 968 oAlaProMetAlaAlaThrGlyLeuProProLeuProAlaGluAsnLysAla----- 986
 Qy 1348 TGCCTATTCTCTCAGTTTGGAGTCTCGAAGCTTGGCTGGGAGCAAAATAATGGCTG 1289
 Db 987 -----PheAlaSerAsnLeuProThrLeuAsnValAlaLysAlaAlaSerSerGly-- 1003
 Qy 1288 CTCTCAAAGTCCCGAGGCTCCATGTAGTGGGTGTGCTCATCAGCATTTTTCGTCTGGA 1229
 Db 1004 -----ProGlyLysPro---SerGlyLeuGlnThrGluSerLysLeu--SerGly 1018
 Qy 1228 CTT-----TCTTCATAGTCTTTTCCAGGCAATGCAT 1199
 Db 1019 LeuLysLysProProThrLeuGlnProSerLysGluAlaLysPheLeuGlu-----His 1036
 Qy 1198 TTGCAAAATGGAAGAACACATGTTTGGCCCTCATAGCACTCACAGTAGTTC-----TTC 1145
 Db 1037 LeuHisLysHisGlnGlySerValLeuHisProAspThrLysThrAlaPheProSerPhe 1056
 Qy 1144 AGCAGCGCTGAGCGCTTACAGTTGCACCCCTTTGCTGTGTCGAAGTTTAGCAGCTCCCA 1085
 Db 1057-GluAsp-----AlaLeuHisArgLeuLeuProThrHisValGlnGlyAla 1072

QY	1313	GTTCCTCAGGACCTCCAAAACTGAGAAAAAATAGGCAGGCGCTTCTCTGTATCTCTCTGG--	1370
DB	360	-----GlyGlyPheHisCysGluCysTrrpva	368
QY	1371	-----GAAGTAGTGGAGGCCACA--	1388
DB	368	lGlyTyrGlnSerSerGlySerLysGluGluAlaCysGluAspValaspGlyCysThrAl	388
QY	1389	-----TGTGCGCTG	1396
DB	388	aAlaTyrSerProCysAlaGlnGlyCysThrAsnThrAspGlySerPheTyrCysSerCy	408
QY	1397	C-----CTGCTGTGCCCGAGGGTCAGGAACGAGCAGCAGGACACTGTCCTCCCAAGCTT	1447
DB	408	sLysGluGlyTyrIleMetSerGlyGluAspSerThrGln-----	421
QY	1448	GGCTGACGAGATGATCTGGAGGAGTTTGGAAAGTGCCCTGTGCGAGATTCCTCCACATCGA	1507
DB	422	-----CysGluAspIleAspGluCysLeuGlyAsn-----	431
QY	1508	GTTCAGTCCCAAGGGCTGAAATTCAGTAGCGTCGACAGCTGCTAAAGGGAATCCCTGT	1567
DB	432	----ProCysAspThrLeuCysIleAsnThrAspGlySerPheArgCysGlyCysProAl	450
QY	1568	GGCA--AGCCTCAGCGCTGGGAATCTGCAC-----CGAGGAAGCTGG-----TG	1609
DB	450	aGlyPheGluLeuAlaProAsnGlyValSerCysThrArgGlySerMetPheSerGluLe	470
QY	1610	CCCAGGAGG-----AGCAGAGCGCGCGCATCA-----	1637
DB	470	uProAlaArgProGlnLysGluAspLysGlyAspGlyLysGluSerThrValProLe	490
QY	1638	-----TGCCAGGTCAGCTGTGAGCTGTGATGATGTCATGTGCTACTGCCAGCCTAC	1690
DB	490	uThrGluMetProGlySerLeuAsnGlySerLysAspValSerAsnArgAlaGlnThrTh	510
QY	1691	T	1691
DB	510	T	510
RESULT 11			
ID	AMFR_HUMAN	STANDARD;	PRT; 323 AA.
AC	P26442;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DE	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Autocrine motility factor receptor precursor (AMF receptor) (GP78).		
GN	AMFR.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91302382; PubMed=1649192;		
RA	Watanabe H., Carmi P., Hogan V., Raz T., Silletti S., Nabi I.R.,		
RA	Raz A.;		
RT	"Purification of human tumor cell autocrine motility factor and		
RT	molecular cloning of its receptor."		
RL	J. Biol. Chem. 266:13442-13448(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=placenta;		
RX	MEDLINE=95352090; PubMed=7626106;		
RA	Huang B., Xie Y., Raz A.;		
RT	"Identification of an upstream region that controls the transcription		
RT	of the human autocrine motility factor receptor."		
RL	Biochem. Biophys. Res. Commun. 212:727-742(1995).		
CC	-!- FUNCTION: SPECIFIC RECEPTOR FOR THE AUTOCRINE MOTILITY FACTOR.		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-!- PTM: PHOSPHORYLATED IN THE PRESENCE OF AMF		

RT ribonucleic acid and of its deduced amino-acid sequence.";
RN Eur. J. Biochem. 147:53-58(1985).
[5]
RP SEQUENCE OF 1-415; 640-737 AND 880-1000 FROM N.A.
RX MEDLINE-88062712; PubMed-3681978;
RA Parma J., Christophe D., Pohl V., Vassart G.;
RT "Structural organization of the 5' region of the thyroglobulin gene.
RT Evidence for intron loss and 'exonization' during evolution.";
J. Mol. Biol. 196:769-779(1987).
[6]
RN SEQUENCE OF 1-25 FROM N.A.
RP MEDLINE-85269632; PubMed-2991855;
RX Christophe D., Cabrer B., Bacolla A., Targovnik H.M., Pohl V.,
RA Vassart G.;
RT "An unusually long poly(purine)-poly(pyrimidine) sequence is located
RT upstream from the human thyroglobulin gene.";
RT Nucleic Acids Res. 13:5127-5144(1985).
[7]
RN SEQUENCE OF 1002-1566 FROM N.A.
RX PubMed-11124863;
RA Moya C.M., Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RT "Genomic organization of the 5' region of the human thyroglobulin
RT gene.";
RN Eur. J. Endocrinol. 143:789-798(2000).
[8]
RN SEQUENCE OF 1645-2768 FROM N.A.
RP MEDLINE-99452200; PubMed-10524569;
RX Mendive F.M., Rivolta C.M., Vassart G., Targovnik H.M.;
RA "Genomic organization of the 3' region of the human thyroglobulin
RT gene.";
RT Thyroid 9:903-912(1999).
[9]
RN PRESENCE OF A 11TH TYROGLOBULIN TYPE-I REPEAT.
RX MEDLINE-96390872; PubMed-8797845;
RA Molina F., Bouanani M., Pau B., Granier C.;
RT "Characterization of the type-I repeat from thyroglobulin, a
RT cysteine-rich module found in proteins from different families.";
RN Eur. J. Biochem. 240:125-133(1996).
[10]
RN IODINATION SITES.
RX MEDLINE-89340430; PubMed-2760035;
RA Lamas L., Anderson P.C., Fox J.W., Dunn J.T.;
RT "Consensus sequences for early iodination and hormonogenesis in human
RT thyroglobulin.";
RN J. Biol. Chem. 264:13541-13545(1989).
[11]
RN SULFATION.
RX MEDLINE-99382264; PubMed-10448091;
RA Niend M.-C., Cauvi D., Venot N., Chabaud O.;
RT "Sulfated tyrosines of thyroglobulin are involved in thyroid hormone
RT synthesis.";
RN Biochem. Biophys. Res. Commun. 262:193-197(1999).
[12]
RN VARIANT GOITER HIS-870.
RX MEDLINE-93164775; PubMed-8094490;
RA Corral J., Martin C., Perez R., Sanchez I., Mories M.T.,
RT San Millan J.L., Miralles J.M., Gonzalez-Sarmiento R.;
RT "Thyroglobulin gene point mutation associated with non-endemic simple
RT goitre.";
RN Lancet 341:462-464(1993).
[13]
RN VARIANTS GOITER R-1264 AND S-1996, AND VARIANTS.
RX MEDLINE-99213884; PubMed-10199792;
RA Hishinuma A., Takamatsu J., Ohshima Y., Yokozawa T., Kuma K.,
RA Yoshida S., Matsura N., Ieiri T.;
RT "Two novel cysteine substitutions (C1263R and C1995S) of thyroglobulin
RT cause a defect in intracellular transport of thyroglobulin in
RT patients with congenital goiter and the variant type of adenomatous
RT goiter.";
RN J. Clin. Endocrinol. Metab. 84:1438-1444(1999).
CC -1- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE
CC (T4) AND TRIIODOTHYRONINE (T3).
CC -1- SUBUNIT: HOMODIMER.

CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: THYROID GLAND SPECIFIC.
CC -1- PTM: SULFATED.
CC -1- DISEASE: DEFECTS IN TG ARE A CAUSE OF FORMS OF GOITER, AN
CC ENLARGEMENT OF THE THYROID GLAND. THIS IS SOMETIMES LINKED WITH
CC HYPOTHYROIDISM.
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYL-ESTERASE/LIPASE FAMILY.
CC -1- SIMILARITY: CONTAINS 11 TYROGLOBULIN TYPE-I DOMAINS.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05615; CAA29104.1; -
CC EMBL; U93033; AAC51924.1; -
CC EMBL; AF230667; AAF62513.2; -
CC EMBL; AF235100; AAG39877.1; -
CC EMBL; AF230666; AAF62515.1; -
CC EMBL; AF305872; AAG17932.1; -
CC EMBL; X02154; CAA26089.1; -
CC EMBL; X06059; CAA29454.1; -
CC EMBL; X06060; CAA29454.1; JOINED.
CC EMBL; X06061; CAA29454.1; JOINED.
CC EMBL; X06062; CAA29454.1; JOINED.
CC EMBL; X06063; CAA29454.1; JOINED.
CC EMBL; X06064; CAA29454.1; JOINED.
CC EMBL; X06065; CAA29454.1; JOINED.
CC EMBL; X06066; CAA29454.1; JOINED.
CC EMBL; X06067; CAA29455.1; -
CC EMBL; X06068; CAA29455.1; JOINED.
CC EMBL; X06069; CAA29456.1; -
CC EMBL; X06870; CAA29456.1; JOINED.
CC EMBL; X02749; CAA26527.1; -
CC EMBL; AF170489; AAD51647.1; -
CC EMBL; AF170486; AAD51647.1; JOINED.
CC EMBL; AF170487; AAD51647.1; JOINED.
CC EMBL; AF170488; AAD51647.1; JOINED.
CC EMBL; AF105687; AAC95473.1; -
CC EMBL; AF105681; AAC95473.1; JOINED.
CC EMBL; AF105682; AAC95473.1; JOINED.
CC EMBL; AF105683; AAC95473.1; JOINED.
CC EMBL; AF105684; AAC95473.1; JOINED.
CC EMBL; AF105685; AAC95473.1; JOINED.
CC EMBL; AF105686; AAC95473.1; JOINED.
CC EMBL; AF080484; AAD50912.2; -
CC EMBL; AF169654; AAD50912.2; JOINED.
CC EMBL; AF169655; AAD50912.2; JOINED.
CC EMBL; AF169656; AAD50912.2; JOINED.
CC EMBL; AF169657; AAD50912.2; JOINED.
CC EMBL; AF169658; AAD50912.2; JOINED.
CC EMBL; AF169659; AAD50912.2; JOINED.
CC EMBL; AF169661; AAD50912.2; JOINED.
CC EMBL; AF169662; AAD50912.2; JOINED.
CC EMBL; AF169663; AAD50912.2; JOINED.
CC EMBL; AF169664; AAD50912.2; JOINED.
CC EMBL; AF080472; AAD50912.2; JOINED.
CC EMBL; AF080473; AAD50912.2; JOINED.
CC EMBL; AF080474; AAD50912.2; JOINED.
CC EMBL; AF080475; AAD50912.2; JOINED.
CC EMBL; AF080476; AAD50912.2; JOINED.
CC EMBL; AF080477; AAD50912.2; JOINED.
CC EMBL; AF080478; AAD50912.2; JOINED.
CC EMBL; AF080479; AAD50912.2; JOINED.
CC EMBL; AF080480; AAD50912.2; JOINED.
CC EMBL; AF080481; AAD50912.2; JOINED.
CC EMBL; AF080482; AAD50912.2; JOINED.
CC EMBL; AF080483; AAD50912.2; JOINED.
CC PIR; A01532; UIHU.
CC PIR; S00014; S00014.

Db 1691 gPheGluProThrGlyPheGlnAsnMetLeuSerGlyLeuTyrAsnProIleValPheSe 1711
 Qy 1767 ---TCYGGGACTCCCTGCC---TGCAATAAAAGAGGGTGAATTTTC 1807
 Db 1711 rAlaSerGlyAlaAsnLeuThrAspAlaHisLeuPheCysLeuAlaCysAspArgAs 1731
 Qy 1808 TACTTGTGTTATGTTGCTTCAATTCCTAGTAGTACCTCCATCAAGTT---AT 1864
 Db 1731. pleuCysCysAspGlyPheValLeuThrGlnValGlnGlyAlaIleIleCysGlyLe 1751
 Qy 1865 TATGAGCCAGCTCAAGTTAGAGCTAGGCTCTTCTTCAGGTGGAGCTCTCCCAAAATCA 1924
 Db 1751 uLeuSerSerProSerValLeu---LeuCysAsnValLysAspTrpMetAspProSerG1 1770
 Qy 1925 CATACAGTCAGGTGGCCATCAGGGGTTTTCAGGCCAGGCCCTGTGACAGGAGATATGG 1984
 Db 1770 u-----AlaTrpAlaAsnAlaThrCysProGlyValThrTyrAspGlnGlu---Se 1786
 Qy 1985 GAGGGGGTGGGTTAGAGCTGGGT 2009
 Db 1786 rHisGlnValIleLeuArgLeuGly 1794

RESULT 13
 CD93_HUMAN
 ID CD93_HUMAN STANDARD; PRT; 652 AA.
 AC Q9NPY3; O00274;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Complement component C1q receptor precursor (Complement component 1, q
 DE subcomponent, receptor 1) (C1qR) (C1qR(p)) (C1q/MBL/SPA receptor)
 DE (CD93 antigen) (CDw93).
 GN C1QR1 OR CD93.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX PubMed=9047234;
 RA Nepomuceno R.R., Henschen-Edman A.H., Burgess W.H., Tenner A.J.;
 RT "cDNA cloning and primary structure analysis of C1qR(P), the human
 RT C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro.";
 RL Immunity 6:119-129(1997).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT ALA-318.
 RX PubMed=11781389;
 RA Steinberger P., Szekeres A., Wille S., Stockl J., Selenko N.,
 RA Prager E., Staffler G., Madic O., Stockinger H., Knapp W.;
 RT "Identification of human CD93 as the phagocytic c1q receptor (C1qRp)
 RT by expression cloning.";
 RL J. Leukoc. Biol. 71:133-140(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead D.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leharvalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownteen R., Sims S.,
 Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 Rogers J.;
 "The DNA sequence and comparative analysis of human chromosome 20.";
 Nature 414:865-871(2001).
 RL [4]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Leukocyte;
 RC Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP CHARACTERIZATION.
 RP PubMed=11994479;
 RA McGreal E.P., Ikewaki N., Akatsu H., Morgan B.P., Gasque P.;
 RT "Human C1qR is identical with CD93 and the mN1-11 antigen but does
 RT not bind C1q.";
 RL J. Immunol. 168:5222-5232(2002).
 RN [6]
 RP O-GLYCOSYLATION.
 RX PubMed=10092817;
 RA Nepomuceno R.R., Ruiz S., Park M., Tenner A.J.;
 RT "C1qR is a heavily O-glycosylated cell surface protein involved in
 RT the regulation of phagocytic activity.";
 RL J. Immunol. 162:3583-3589(1999).
 CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
 CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant
 CC protein A (SPA). May mediate the enhancement of phagocytosis in
 CC monocytes and macrophages upon interaction with soluble defense
 CC collagens. May play a role in intercellular adhesion.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Highly expressed in endothelial cells,
 CC platelets, cells of myeloid origin, such as monocytes and
 CC neutrophils. Not expressed in cells of lymphoid origin.
 CC -!- PTM: N- and O-glycosylated.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
 CC -!- CAUTION: Has been sometimes referred to as a collectin receptor.
 CC -!- CAUTION: According to Ref.5, C1q is not a ligand for C1QR1.
 CC -!- DATABASE: NAME=PROW; NOTE=PROW 3:1-6(2001);
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/467246456.g.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U94333; AAB53110.1;
 CC EMBL; AL118508; CAG00597.1;
 CC EMBL; BC028075; AAK28075.1;
 CC HSSP; P35555; 1EMN.
 CC MIM; 120377;
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR001881; EGF-Ca.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR001187; Tissue_factor.
 CC Pfam; PF00008; EGF; 5.
 CC Pfam; PF00059; lectin_C; 1.
 CC Pfam; PF01108; Tissue_fac; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF_CA; 3.
 CC SMART; SM00001; EGF_like; 2.
 CC PROSITE; PS00610; ASX_HYDROXYL; 3.
 CC PROSITE; PS00615; C_TYPE_LLECTIN_1; FALSE_NEG.
 CC PROSITE; PS00641; C_TYPE_LLECTIN_2; 1.
 CC PROSITE; PS01186; EGF_2; 3.

DR PROSITE; PS01187; EGF_CA: 3.
 KW Receptor; EGF-like domain; Signal; Transmembrane; Glycoprotein;
 KW Repeat; Lectin; Polymorphism.
 FT SIGNAL 1 21
 FT CHAIN 22 652
 FT DOMAIN 24 580
 FT TRANSMEM 581 601
 FT DOMAIN 602 652
 FT DOMAIN 32 174
 FT DOMAIN 260 301
 FT DOMAIN 302 344
 FT DOMAIN 345 384
 FT DOMAIN 385 426
 FT DOMAIN 427 468
 FT DOMAIN 594 601
 FT DISULFID 264 275
 FT DISULFID 271 285
 FT DISULFID 287 300
 FT DISULFID 306 317
 FT DISULFID 311 328
 FT DISULFID 330 343
 FT DISULFID 349 358
 FT DISULFID 354 367
 FT DISULFID 369 383
 FT DISULFID 389 400
 FT DISULFID 396 409
 FT DISULFID 411 425
 FT DISULFID 431 443
 FT DISULFID 439 452
 FT DISULFID 454 467
 FT CARBOHYD 325 325
 FT VARIANT 318 318
 FT CONFLICT 22 22
 FT CONFLICT 36 36
 FT CONFLICT 38 39
 FT CONFLICT 155 155
 FT CONFLICT 186 186
 FT CONFLICT 492 492
 FT CONFLICT 496 496
 FT CONFLICT 504 504
 FT CONFLICT 541 541
 SQ SEQUENCE 652 AA; 68560 MW; EECADFAC55FCAC2 CRC64;
 Alignment Scores:
 Pred. No.: 0.633 Length: 652
 Score: 112.00 Matches: 129
 Percent Similarity: 27.11% Conservative: 54
 Best Local Similarity: 19.11% Mismatches: 183
 Query Match: 2.79% Indels: 309
 DB: 1 Gaps: 30
 US-09-743-237-1 (1-2241) x CD93_HUMAN (1-652)
 QY 39 GGCTCAGGACCAACAGGGGATACACAG-----TGTGTGTTCTGGGCTGTGGACT 89
 DB 19 GlyAlaCylThrGlyAlaAspThrGluAlaValValCysValGlyThrAlaCysThr 38
 QY 90 TGTGACTCCACCACTCCGCCCCAGCAGGCT----- 122
 DB 39 AlaHisSerGlyLysLeuSerAlaAlaGluAlaGlnAsnHisCysAsnGlnAsnGlyGly 58
 QY 123 -----AGGATAGAACCCAGGCCCTTTGGTGTCTG-----CAG 158
 DB 59 AsnLeuAlaThrValLysSerLysGluAlaGlnHisValGlnArgValLeuAlaGln 78
 QY 159 ATAGTC----- 164
 DB 79 LeuLeuArgArgGluAlaAlaLeuThrAlaArgMetSerLysPheTrpIleGlyLeuGln 98
 QY 165 -----TTCAGCTGTGTAGTTGGGTTGGCTGGGA 194
 DB 99 ArgGluLysGlyLysCysLeuAspProSerLeuProLeuLysGlyPheSerTrpValGly 118

QY 195 -----GATTTTTTTTCTTCCACACCAAGAGCTTCCATTATTGAGGATTTTTCAGT 245
 DB 119 GlyGlyGluAsp-----ThrProTyrSerAsnTrpHisLysGlu----- 131
 QY 246 TGATGATCTCCCTCTGTGAAGATAGGACAGTCTCTTTAACTATGTAGAGTTTGA 305
 DB 131 ----- 131
 QY 306 TGAATTCGTCTTCAACCATATTGCTAAGCTATATAGCAATTCCTTGAATTCGTATAT 365
 DB 132 -----LeuArgAsnSerCysIleSerLys-----ArgCysValSerLeuLeuLeu 146
 QY 366 AACTTAGGAGAACCTGATTCTCTGCTCTCATCATCTGAGTGTAGTGTACAGGGGG 425
 DB 147 AspLeuSerGlnProLeuLeuProSerArgLeu---ProLys----- 159
 QY 426 AATCATTTTGGTGAGACTCGGATGAAGTACTGCTCCAGGTTCCCAAGGACGACCAAGCA 485
 DB 160 -----TrpSerGluGlyProCysGlySerProGlySerProGlySer----- 173
 QY 486 AGAAAAAGTGTGAATCAAAAGACAGGTTGTAGTGTGCCAGC----- 530
 DB 174 -----AsnIleGluGlyPheValCysLysPheSerPheLysGlyMetCysArgProLeu 191
 QY 531 -----GGCAGCCCTGAAGACGCGACTTTCAGGCCCTCTGGCTCAGGAATCCTGTTC 584
 DB 192 AlaLeuGlyGlyProGlyGlnValThrTyrThrProPheGlnThrThrSer----- 209
 QY 585 AAGTTCCTCATCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 626
 DB 210 -----SerSerLeuGluAlaValProPheAlaSerAlaAlaAsnValAlaCysGly 226
 QY 627 -----CCTCGAAGAAAGACTCCAGCCCATGGTGTGTTGTGCTGAGCTGAA----- 671
 DB 227 GluGlyAspLysAspGluThrGlnSerHisTyrPheLeuCysLysGluLysAlaProAsp 246
 QY 672 -----GGAGCGCCAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719
 DB 247 ValPheAspTrpGlySerSerGlyProLeuCysValSer----- 259
 QY 720 AAAGCGCTCCATCTGCTTCTCTCAGTACGATGACAGAGCAGTTCCTCCTCAGTACAGCTC 779
 DB 260 -----ProLysTyr----- 262
 QY 780 CCTAAGCCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAGGAAAGTTAAATPCTC 839
 DB 262 ----- 262
 QY 840 ATCACACAGGTTGATAATGAGCTCTCCCATCAGCTGTCAATGGGGTGCCTTTCCTCTCT 899
 DB 263 -----GlyCysAsnPheAsnAsn 268
 QY 900 GGACCTGCTCTGCAAGGGCCCAACCAATAACTCTCTCTGGGTACTGTGCTGCTCTCTCT 959
 DB 269 GlyGlyCysHisGln-----AspCysPheGlu 277
 QY 960 AGCGGGGACTTCTGCAACAGCTGCAACAACTGCGCCATGAGCTCGAGCGCTTC 1019
 DB 278 GlyLysAspGlySerPheLeuCysGlyCys-----ArgProGlyPheArgLeuLeu 294
 QY 1020 AAGCCATAAAGCGGTCTTGTATGAATTCCTGAAGCTTTTCCAAACCAAAATGGGAAA 1079
 DB 295 AspAspLeuValThrCysAlaSerArgAsnProCysSerSerSerProCysArgGly--- 313
 QY 1080 GGCGCTGGGAGCTCTAACTTCCACACAGCAAGGGTGCACACTGTAAAGCCTCAGGC 1139
 DB 314 -----GlyAlaThrCysValLeuGlyPro 321
 QY 1140 TGCTTGAAGAACTAC---TCTGAGTGTCTATGAGGCCCAAAATCATGTGTTCTTCATTTGC 1196
 DB 322 HisGlyLysAsnTyrThrCysArgCysProGlnGlyTyrGlnLeuAspSerSerGlnLeu 341


```
QY 845 ACAGGTGATGATGAGCTCTCCCATCAGCTGCTCAATGGGGTGGCTTTCCTCTGGACC 904
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 135 lnaArgLeuGlnAsnValGlnValGlnAsnGlnLysValGlyMet---PheGluAlaHisI 154
QY 905 TGCCTGCAAGGGCCACCCCAAAATAACTCTGCTGGGACTGTGACTGCTCTCCAGCGG 964
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 154 leGlnAlaGlnSer-----SerAlaIleGlnAlaProArgSerP 167
QY 965 GCACITCTCCACAGCTGCAGCTGCAACACACCTGCGCCATGAGCTCGAGCGTTCAAAGC 1024
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 167 roArgLeuGlyArgAlaArgSerProCysProPheArgSerSer---SerGlnP 186
QY 1025 CATAAAGGCGTCTTGTATAGAAATCCTGAAGCTTTCACCAACCAAAAA-----TGGG 1075
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 186 roProGlyArgValLeuValGlnGlnAlaArgSerGluGluArgArgThrLysSerTrpG 206
QY 1076 GAAAGCCCTCGGAGCTGTAAACTTTCGACACAGCA----- 1113
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 206 lylGluGlnCysProGluThrSerGlyThrAspSerGlyArgLysGlyProSerLeuC 226
QY 1114 -----AAGGTGCACTGTAAAGCGCTCAGGCTCGCTGAAGAACTACTG 1156
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 226 ysSerSerGlnValLysLysGlyMetProProLeuProGlyArgAlaProThrGlyS 246
QY 1157 TGAGTGTATGAGGCCAAATCATGTGT-----CTTCATTTCGCAAAATG 1201
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 246 erGluAlaGlnGlyProSerAlaPheValArgMetGluLysGlyIleProAlaSerProA 266
QY 1202 CATTCCTGCAAAACTATGAAGAAGTCCAGAACCAAAATGCTGTAGACACACCCCA 1261
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 266 rgCysGlySerProThrAlaMetGluIleAspLysArg-----GlySerProT 282
QY 1262 CTACATGG-----AGCCTGGGACTTTTGAGACGACCAT----- 1296
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 282 hrProGlyThrArgSerCysLeuAlaProSerLeuGlyLeuPheGlyAlaSerLeuThrM 302
QY 1297 -----ATTGTCCCCAGCCAAAGTTCT-----CAGGACCTCC 1327
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 302 etAlaThrGluValAlaAlaArgValThrSerThrGlyProHisArgProGlnAspLeuA 322
QY 1328 AAACTGAGAAAAATAGCCAG-----CCTTCCTCTGTAT 1363
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 322 laLeuThrGluProSerGlyArgAlaArgGluLeuGluAspLeuGlnProProGluAla- 341
QY 1364 CTCCTGGGAAGTAGTGA-----GGCCACATGTGCTCGCTGGCCCGCCAGGTGA 1414
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 342 LeuValGluArgGlnGlyGlnPheLeuGlySerGluThrSerProAlaProGluArgGly 361
QY 1415 GGAAGCAGACGAGGACACTGTTCCTCCCAAGCTTGCTGAGCAGATGATCTGGAGGAGTT 1474
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 362 GlyProArgAspGlyGluProProGlyLysMetGlyLysGlyTyLeuProCysGly--- 380
QY 1475 TGAAGGTGCCTGTCGAGATCTCCACAT-----CGAGTT 1510
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 381 -----MetProGlySerGlyGluProGluValGlyLysArgProGluGluThrVal 398
QY 1511 CAAGTCCAAAGGGCTGAAAATTGACTAGCTGCAAGCTGGTAAAGGGGAATGCTGTGGC 1570
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 399 SerValGlnSerAlaGlnSerSerAspAla-LeuSerTrpSerArg---LeuProArgAl 417
QY 1571 AAGCTCAGCCCTGGGAATCTGCACCGAGGAAGCTGGTCCCGGAGGAGGACGAGCGCG 1630
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 417 aLeuAlaSerValGly-----ProGluGluAlaArgSerGl 429
QY 1631 CCATCATGCGGCAGCTCAGCTGTGAGTCTGAGTGTGATCTGATGTGCTACTGGCCGCCCTAC 1690
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 429 yAlaProValGlyGlyArgTrpGlnLeuSerAspArgValGluGlyLysProTh 449
QY 1691 TCAAGGTATCTCTA-----AAGTGAACGAGCAGGACCCCTGGGGATGGACACTGGCCCT 1747
   |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 449 rLeuGlyLeuLeuGlyGlySerProSerAlaGlnProGlyThrGly-----As 465
```

```
QY 1748 CCTGTCCTGGGAGGCGCTCTGGG-----GACTCCTGCTCCCTGCATAAAAGAGG 1798
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 465 nValGluAlaGlyIleProSerGlyArgMetLeuGluProLeuProCys----- 481
QY 1799 GTGATTTTCTACTTCTGTTATGTCTTTGCTTTCAAAATGCTTAGTACCTCATTCAC 1858
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 482 -----TrpAspAlaAl 485
QY 1859 AGTTATTATGACCCAGCCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGGACTGTGCC 1918
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 485 aLysAspLeuLysGluProGln-----CysPr 494
QY 1919 AAATCACATACAAGTCAGGTGGCCATCAGGGGTTTTTCCAGGCCAGGCTGTGCACAGGAG 1978
   |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 494 oProGlyAspArgVal-----GlyValGlnProGlyAsnSer-----Arg-- 507
QY 1979 ATATGGGAGGGGCTGGGTAGAGCTGGTGTGTTGGT 2017
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 508 ValTrpGlnGlyThrMetGluLysAlaGlyLeuAlaTrp 520
RESULT 15
OC90_HUMAN STANDARD: PRT; 493 AA.
AC 002509;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN Otcocin 90 precursor (OC90) (Phospholipase A2 homolog).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9311166; PubMed=8382789;
RA Feuchter-Murthy A.E., Freeman J.D., Mager D.L.;
RT "Splicing of a human endogenous retrovirus to a novel phospholipase
RL Nucleic Acids Res. 21:135-143(1993).
CC -1- FUNCTION: IT IS UNLIKELY THAT THIS PROTEIN HAS PHOSPHOLIPASE A2
CC ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. CONTAINS 3
CC PA2 TYPE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: Z14310; CAA78662.1; ALT_INIT.
DR HSSP; P00593; 4BP2.
DR Genew; HGNC:8100; OC90.
DR MIM; 601658;
DR InterPro; IPR001211; PhospholipaseA2.
DR Pfam; PF00068; Phoslip; 2.
DR PRINTS; PR00389; PHPLIPASEA2.
DR PRODOM; PD000303; PhospholipaseA2; 2.
DR SMART; SM00085; PA2C; 2.
DR PROSITE; PS00118; PA2_HIS; 2.
DR PROSITE; PS00119; PA2_ASP; 1.
DR Signal; Glycoprotein; Repeat.
KW SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 493 OTOCININ 90.
FT DOMAIN 76 190 PHOSPHOLIPASE A2-LIKE 1.
FT DOMAIN 321 377 PHOSPHOLIPASE A2-LIKE 2.
FT DOMAIN 389 441 PHOSPHOLIPASE A2-LIKE 3.
FT DISULFID 85 145 BY SIMILARITY.
FT DISULFID 99 190 BY SIMILARITY.
```

```
FT DISULFID 101 117 BY SIMILARITY.
FT DISULFID 116 172 BY SIMILARITY.
FT DISULFID 123 165 BY SIMILARITY.
FT DISULFID 132 158 BY SIMILARITY.
FT DISULFID 152 163 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 493 AA; 53338 MW; 59F9AF7C1364A5B7 CRC64;

Alignment Scores:
Pred. No.: 0.781 Length: 493
Score: 110.50 Matches: 86
Percent Similarity: 33.26% Conservative: 58
Best Local Similarity: 19.86% Mismatches: 135
Query Match: 2.76% Indels: 154
DB: 1 Gaps: 21

US-09-743-237-1 (1-2241) x OC90_HUMAN (1-493)
QY 531 GGCAGGCCCTGAACACGAGCTTTCCAGGCCCTCTGGCTCAGGAATCTCTTGCAGTTC 590
Db 107 GlyLeuProValAspGluSer-----AspSerCysCys---Phe 118
QY 591 CCATCATCCAGGAGGAGGAGGCTCCAGCTGCCCTCGGAGAAAGACTCCAGGCC 650
Db 119 GlnHisArgCysCysGluGluAlaAlaGluMetAspCysLeuGlnAspProAlaLys 138
QY 651 ATG-----GTGATTGTGCTGAGTGAAGAGGCGCC 680
Db 139 LeuSerThrGluValAspCysValGlyLysLysLysLysLysLysLysLysLys 158
QY 681 CAG---ATGCTCTGCATAGACACTGTGGCGCGAGGAGGCTCAAGCG----- 725
Db 159 GluHisLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 178
QY 726 -----CTCATCTGCTCTCTCAGTACGATGACGAGCAGAGTTCCTCAGTCAGAG 776
Db 179 AsnSerSerLeuAsnLeuLeuAspThrSerPheCysLeuAlaGlnThrProGluThr 198
QY 777 CTCCCTAAGCCATGACAACTTAGTGGGAAGACTTCTGCCAGTA---CCAGCGAAGTTA 833
Db 199 IleLysGluAspLeuThrThrLeuLeuProArgValValProValGluProThrAsp 218
QY 834 AATCTCATC-----ACACAGCTTGAT----- 854
Db 219 SerLeuThrAlaLeuSerGlyGluValAlaAlaGluThrGluAlaAspArgLeuLeu 238
QY 855 -----AATGGAGCTCTCCCATCAGCT 875
Db 239 LeuSerLysLysLysAlaGlyHisAspGlnGluGlyValGlyAlaAlaArgAlaThr 258
QY 876 GTCAATGGGCTGCC-----TTT 893
Db 259 ProGlySerAlaGluLeuValAlaThrArgValThrAlaLysIleValThrLeuVal 278
QY 894 CCCTCTGGA-----CCTGCTCTGCAAGGCCACCCAAATA 929
Db 279 ProAlaGlyIleLysSerLeuGlyLeuAlaValSerValGluAsnAspProGluGlu 298
QY 930 ACTCTGCTGGTACTGCTACTCTCTCCAGCGGGGACTTCTGCAACAGCTGCAGCTGC 989
Db 299 ThrThrGluLysAlaCysAspArgPheThr-----PheLeuHisLeuGlySerGly 315
QY 990 AACAACTGGCCCATGAGCTGAGCGCTTCAAGCCATAAAGCGGTCTTGTATAGAAAT 1049
Db 316 AspAsnMetGlnValMetProGlnLeuGlyGluMetLeuPheCysLeuThrSerArgCys 335
QY 1050 CCTGAAGCTTTCACCAAAATGGGGAAGGCGCTGCGGAGCTGCTAAACTTCGACAC 1109
Db 336 ProGluGluPheGlu----- 340
QY 1110 AGCAAAAGGTGCAACTGTAAGCGCTCAGGC----- 1139
```

```
Db 341 SerTyrGlyCysTyrCysGlyGlnGluGlyArgGlyGluProArgAspAspLeuAspArg 360
QY 1140 ---TGCTGAAGAACTACTGTGAGTGTCTATGAG----- 1169
Db 361 CysCysLeuSerHisCys---CysLeuGluGlnValArgArgLeuGlyCysLeuLeu 379
QY 1170 -----GCCAAATCATGTGT-----TCT 1187
Db 380 GluArgLeuProTrpSerProValValCysValAspHisThrProLysCysGlyGln 399
QY 1188 TCCATTTCGAAATGTCATTGCTTGCAGAACTATCAAGAAATCCAGAAACGAAAAATGCTG 1247
Db 400 SerLeuCysGluLysLeuLeuCysAlaCysAspGlnThrAlaAlaGlu----- 415
QY 1248 ATGAGCACACCCCACTACATGGAGCGCTGGGACTTTGAGAGCAGCCATTATTTGTCGCCA 1307
Db 416 -----CysMetThrSerAlaSerPheAsnGlnSer----- 425
QY 1308 GCCAAGTTCTCAGAGCTCCAAACTCAGAAAAAATAGGAGCGCTTCTCTCTATCTCC 1367
Db 426 -----LeuLysSerProSerArgLeuGlyCysProGlyGlnProAlaAlaCysGluAsp 443
QY 1368 TGGGAAGTAGTGGAGGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1427
Db 444 ---SerLeuHisProValProAlaAlaProThrLeuGlySerSerSerGluGluAspSer 462
QY 1428 GAGCACTGTTCCTCAAGCTTTGGCTGAGCAGATGATCTGAGGAGGTTTGGAGGTGCTG 1487
Db 463 GluGluAspProProGln-----GluAspLeuGlyArg-Ala-- 474
QY 1488 TCCAGATTTCTCCACATCGAGTTCAGTCCCAAGGGGC 1524
Db 475 -LysArgPheLeuArgLysSerLeuGlyProLeuGly 486

Search completed: April 21, 2003, 11:29:51
Job time : 64.0475 secs
```


GenCore version 5.1.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:20:22 ; Search time 141.77 Seconds
(without alignments)
5514.106 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 4008

Sequence: 1 tatctctgtgggtggccgcg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame_n2p.model -DEV=rlp
-Q/cgn2_1/USPTO.spool/US0743237/runat_21042003_111944_399/app_query.fasta_1.4494
-DB-SPTREMBL_21 -QFMT-fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=-1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0743237@cgn_1.1.462@runat_21042003_111944_399 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1881.5	46.9	475	11	Q8VIE0	Q8vie0 mus musculus

2	1871.5	46.7	475	11	Q8VIE1	Q8vie1 mus musculus
3	786.5	19.6	277	11	Q9D571	Q9d571 mus musculus
4	440.5	11.0	950	5	Q9V6Q8	Q9v6q8 drosophila
5	400.5	10.0	429	5	O62295	O62295 caenorhabdi
6	400	10.0	435	5	Q95QD7	Q95qd7 caenorhabdi
7	371	9.3	571	10	Q9SL70	Q9sl70 arabidopsis
8	366	9.1	603	10	Q9SZD1	Q9szd1 arabidopsis
9	266.5	6.6	601	10	Q9CAV1	Q9cav1 arabidopsis
10	253.5	6.3	356	10	Q9LW71	Q9lw71 arabidopsis
11	248.5	6.2	896	10	O23333	O23333 arabidopsis
12	246.5	6.2	526	10	Q9ZS22	Q9zs22 glycine max
13	228	5.7	526	10	Q94AI2	Q94ai2 arabidopsis
14	228	5.7	695	10	Q9LE32	Q9le32 arabidopsis
15	228	5.7	695	10	Q9LUI3	Q9lui3 arabidopsis
16	226.5	5.7	609	10	Q9M679	Q9m679 arabidopsis
17	186	4.6	243	5	Q9VMQ3	Q9vmq3 drosophila
18	183.5	4.6	593	10	Q9LUI5	Q9lui5 arabidopsis
19	168.5	4.2	553	10	Q94DS2	Q94ds2 oryza sativ
20	150.5	3.7	1175	11	P70200	P70200 mus musculus
21	146.5	3.6	1175	11	O35126	O35126 mus musculus
22	146	3.6	1737	4	O75097	O75097 homo sapien
23	131.5	3.3	1428	5	O44341	O44341 hallotis ru
24	131.5	3.3	4123	4	O75851	O75851 homo sapien
25	126.5	3.2	1481	4	Q9ULI4	Q9uli4 homo sapien
26	125.5	3.1	1182	4	Q99495	Q99495 homo sapien
27	124	3.1	3680	5	O9V808	O9v808 drosophila
28	123.5	3.0	1235	4	O95428	O95428 homo sapien
29	123.5	3.0	4957	4	O14687	O14687 homo sapien
30	123.5	3.0	5262	4	O14686	O14686 homo sapien
31	123	3.0	1190	4	Q99621	Q99621 homo sapien
32	122.5	3.1	1130	4	Q9HIV5	Q9hiv5 homo sapien
33	121	3.0	955	4	Q96DN2	Q96dn2 homo sapien
34	120.5	3.0	5146	6	Q8SPM4	Q8spm4 bos taurus
35	119.5	3.0	975	10	Q9MIX1	Q9mix1 arabidopsis
36	119.5	3.0	2843	4	Q9Y6R7	Q9y6r7 homo sapien
37	118.5	3.0	610	6	Q95LGI	Q95lgi equus cabal
38	117.5	2.9	1574	11	O88281	O88281 rattus norv
39	117.5	2.9	2971	4	Q9V5L9	Q9v5l9 homo sapien
40	116	2.9	383	5	O9Y075	O9y075 leishmania
41	116	2.9	544	10	Q9LWU0	Q9lwu0 oryza sativ
42	115.5	2.9	938	11	Q9JL72	Q9jl72 mus musculus
43	115	2.9	1698	5	Q94438	Q94438 chironomus
44	115	2.8	4823	13	O93321	O93321 fugu rubrip
45	114.5	2.9	1151	13	O57580	O57580 gallus gall

ALIGNMENTS

RESULT 1
Q8VIE0 PRELIMINARY; PRT; 475 AA.
AC Q8VIE0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Tesmin (Hypothetical 50.6 kda protein).
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
protein; a possible multifunctional protein with dynamic changes of
localization throughout spermatogenesis."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; AB057423; BAB64935.1; -
DR EMBL; BC024377; AAH24377.1; -
DR InterPro; IPR005172; CXG.
DR Pfam; PF03638; CXG; 1.
KW Hypothetical protein.
SQ SEQUENCE 475 AA; 50615 MW; 4373ED1188DCF79F CRC64;

Alignment Scores:
Pred. No.: 5,76e-173 Length: 475
Score: 1871.50 Matches: 353
Percent Similarity: 96.47% Conservative: 2
Best Local Similarity: 95.92% Mismatches: 6
Query Match: 46.94% Indels: 7
DB: 11 Gaps: 1

US-09-743-237-1 (1-2241) x Q8VIE0 (1-475)
QY 453 CTACTGCCAGGTTCACAGGAGCAGCAAGCAAGCAAGCAAGT----- 494
Db 108 LeuLeuProGlyAlaArgGlyProAlaLeuLeuProLeuSerAlaGlyValArgValIle 127
QY 495 ---GTTGAAATCAAGAAGCAGGTGGTAGTGCCAGCGCGGAGCAGCAGT 551
Db 128 ProValGluIleLysGluAlaGlySerValProGlyGlySerProGluAspAla 147
QY 552 TTCCAGGCCCTCTGGCTCAGGAATCCTGTGCAAGTTCCCATCATCCAGGAGCAGAG 611
Db 148 PheGlnAlaProLeuAlaGlnSerCysLysPheProSerSerGlnGluAlaGlu 167
QY 612 GAGGCTCCAGCTGCCCTCGGGAAGAAAGACTCCAGCCCATCGGTGATTGTGAGTGA 671
Db 168 GluAlaSerSerCysProArgLysLysAspSerSerProMetValIleCysGlnLeuLys 187
QY 672 GGAGCGCCAGACTCTGCTGATACACACTGTGCGCGAGGAGCTCAAGCGCTCCAT 731
Db 188 GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis 207
QY 732 CTGCTTCTCTCAGTACGACGAGCAGTTCCTCCTCAGTACGAGCTCCCTAAGCAATG 791
Db 208 LeuLeuProGlnTyrAspAspGlnSerSerPheProGlnSerGluLeuProLysProMet 227
QY 792 ACACTTTAGTGGGGAAGACTTCTGCCAGTACCAGGAGAAAGTAAATCTCATCACAGGTT 851
Db 228 ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal 247
QY 852 GATAATGAGCTCTCCCATCAGCTCTCAATGGGGCTGCTTCCCTCTGGACCTGCTCTG 911
Db 248 AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu 267
QY 912 CAAGGCCACCCAAATAACTCTGTCTGGTACTGTGACTGCTTCCAGCGGGGACTTC 971
Db 268 GlnGlyProProLysIleThrLeuSerGlyTyrCysAspCysPheSerSerGlyAspPhe 287
QY 972 TCCACAGCTGAGCTGCAACACTCGCCCATGAGCTGAGCTGAGCGCTTCAAGCCATAAG 1031
Db 288 CysAsnSerCysSerCysAsnLeuArgHisGluLeuLeuArgPheLysAlaIleLys 307
QY 1032 GGTGCTCTGTAGAAATCTTGAAGCTTTCACCAAAATGGGAAAGCGCTGGGA 1091
Db 308 AlaCysLeuAspArgAsnProGluAlaPheGlnProLysMetGlyLysGlyArgLeuGly 327
QY 1092 GCTGCTAACTTCGACAGCAAAAGGTCGAACCTGTAAGCGCTCAGCGTGGCTGAAGAAC 1151
Db 328 AlaAlaLysLeuArgHisSerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsn 347
QY 1152 TACTGTGAGTGTATGAGCCAAATCATGTGTTCTTCCATTTGCAATGATGCTTGC 1211
Db 348 TyrCysGluCysTyrGluAlaLysIleMetCysSerSerIleCysLysCysIleAlaCys 367
QY 1212 AAAAATATGAAGAAGTCCAGAACGAAATGCTGATGAGCACACACCCCTACATGAG 1271
Db 368 LysAsnTyrGluGluSerProGluArgLysMetLeuMetSerThrProHisTyrMetGlu 387

RESULT 2
Q8VIE1 PRELIMINARY; PRT; 475 AA.
ID Q8VIE1
AC Q8VIE1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-JUN-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tesmin.
GN TESM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Sutou S., Miwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;
RT "Structure of the tesmin gene encoding a testis-specific persistent
RT protein: a possible multifunctional protein with dynamic changes of
RT localization throughout spermatogenesis.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB057422; BAB64934.1; -
DR InterPro; IPR005172; CXG.
DR Pfam; PF03638; CXG; 1.
SQ SEQUENCE 475 AA; 50633 MW; 4373F3DEA27393AF CRC64;

Alignment Scores:
Pred. No.: 5,35e-172 Length: 475
Score: 1871.50 Matches: 351
Percent Similarity: 95.92% Conservative: 2
Best Local Similarity: 95.38% Mismatches: 8
Query Match: 46.69% Indels: 7
DB: 11 Gaps: 1

US-09-743-237-1 (1-2241) x Q8VIE1 (1-475)
QY 453 CTACTGCCAGGTTCACAGGAGCAGCAAGCAAGCAAGCAAGT----- 494
Db 108 LeuLeuProGlyAlaArgGlyProAlaLeuLeuProLeuSerAlaGlyValArgValIle 127
QY 495 ---GTTGAAATCAAGAAGCAGGTGGTAGTGCCAGCGCGGAGCAGCAGT 551
Db 128 ProValGluIleLysGluAlaGlySerValProGlyGlySerProGluAspAla 147
QY 552 TTCCAGGCCCTCTGGCTCAGGAATCCTGTGCAAGTTCCCATCATCCAGGAGCAGAG 611
Db 148 PheGlnAlaProLeuAlaGlnSerCysLysPheProSerSerGlnGluAlaGlu 167
QY 612 GAGGCTCCAGCTGCCCTCGGGAAGAAAGACTCCAGCCCATCGGTGATTGTGAGTGA 671
Db 168 GluAlaSerSerCysProArgLysLysAspSerSerProMetValIleCysGlnLeuLys 187
QY 672 GGAGCGCCAGACTCTGCTGATACACACTGTGCGCGAGGAGCTCAAGCGCTCCAT 731
```

```
Db 188 GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis 207
QY 732 CTGCTTCTCAGTACGATGACAGACAGAGTTCCTCCAGTACAGAGTCCCTAAAGCCAAATG 791
Db 208 LeuLeuProGlnTyrAspAspGlnSerSerPheProGlnSerGluLeuProLysProMet 227
QY 792 ACAACTTTAGTGGAGAGCTTCGCCAGTACACGAGGAGTAAATCTCATCACACAGGTT 851
Db 228 ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal 247
QY 852 GATAATGGAGCTCTCCATCAGCTGCAATGGGGCTGCCCTTCCCTCTGGACCTGCTCTG 911
Db 248 AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu 267
QY 912 CAAGGGCCACCCAAATAACTCTGCTGGTACTGTACTGTCTTCCAGCGGGGACTTC 971
Db 268 GlnGlyProProLysIleThrLeuSerGlyTyrCysAspCysPheSerSerGlyAspPhe 287
QY 972 TGCAACAGCTGCAGCTGCAACACCTGCGCCATGAGCTCGAGGCTTCAAGGCCATAAAG 1031
Db 288 CysAsnSerCysSerCysAsnAsnLeuArgHisGluLeuGluArgPheLysAlaIleLys 307
QY 1032 GCGTGTCTGTATGAATACTGCTGGTACTGTACTGTCTTCCAGCGGGGACTTC 1091
Db 308 AlaCysLeuAspArgAsnProGluAlaPheGlnProLysMetGlyLysGlyArgLeuGly 327
QY 1092 GCTGCTAAATTCGACACAGCAAGGTCGAACCTGTAAAGGCTCAGGCTGCCCTGAGAAC 1151
Db 328 AlaAlaLysLeuArgHisSerLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsn 347
QY 1152 TACTGTGAGTGCATGAGGCCAAATCATGTGTTCTTCCATTTGCCAATGCAATGCTTGC 1211
Db 348 TyrCysGluCysTyrGluAlaLysIleMetCysSerSerIleCysLysCysIleAlaCys 367
QY 1212 AAAAATATGAAGAAGTCCAGAACGAAATAATGCTGATGAGCACACCCCACTACATGGAG 1271
Db 368 LysAsnTyrGluGluSerProGluArgLysMetLeuMetSerThrProHisTyrMetGlu 387
QY 1272 CTTGGGGACTTTCAGAGCAGCAATTTGCTCCCGCCAGGCTTCCTCAGGACCTCCAAA 1331
Db 388 ProGlyAspPheGluSerHisTyrLeuSerProAlaLysPheSerGlyProProLys 407
QY 1332 CTGAGAAAATAAGCAGGCTTCTCTGTATCTCTGGAAGTGTAGGAGGCGCACATGT 1391
Db 408 LeuArgLysAsnArgGlnAlaPheSerCysIleSerTrpGluValGluAlaThrCys 427
QY 1392 GCTGCTGCTGCTGCCAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1451
Db 428 AlaCysLeuLeuAlaHisGlyGluGluAlaGluHisGluHisCysSerProSerLeuAla 447
QY 1452 GAGCAGATGATCTCGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAGGAGTGGAG 1511
Db 448 GluGlnMetIleLeuGluGluPheGlyArgCysLeuSerGlnIleLeuHisIleGluPhe 467
QY 1512 AAGTCCAAAGGGCTGAAAATTGAG 1535
Db 468 LysSerLysGlyLeuLysIleGlu 475
```

RESULT 3

```
Q9D571 ID Q9D571 PRELIMINARY; PRT; 277 AA.
AC Q9D571;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Adult male testis cDNA, RIKEN full-length enriched library,
DE clone:4930509C02, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Frieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015732; BAB29949.1; -
SQ SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;
```

Alignment Scores:

```
Pred. No.: 4,52e-67 Length: 277
Score: 786.50 Matches: 155
Percent Similarity: 92.35% Conservative: 2
Best Local Similarity: 91.18% Mismatches: 6
Query Match: 19.62% Indels: 7
DB: 11 Gaps: 1
```

US-09-743-237-1 (1-2241) x Q9D571 (1-277)

```
QY 453 CTACTGCCAGTTCCCAAGGCGAGCAAGCAAGCAAGCAAGCAAGT----- 494
Db 108 LeuLeuProGlyAlaArgGlyProAlaLeuLeuLeuProLeuSerAlaGlyValArgValIle 127
QY 495 ---GTTGAATAACAAGAGCAGGTGTGTGTCAGAGGCGGCGCCCTGAAGCGAGCT 551
Db 128 ProValGluIleLysGluAlaGlyGlySerValProGlyGlySerProGluAspAlaAla 147
QY 552 TTCCAGGCGCCCTCGCTCAGGAATCTCTTGGCAAGTTCCTCATCATCCAGGAGGAGCAG 611
Db 148 PheGlnAlaProLeuAlaGlnGluSerCysLysPheProSerSerGlnGluAlaGlu 167
QY 612 GAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGGCGCCCATGTTGTTCTCAGCTGAAA 671
Db 168 GluAlaSerSerCysProArgLysLysAspSerSerPrometValIleCysGlnLeuLys 187
QY 672 GGAGGCGCCAGATGCTGTCATAGACAACTGTGGCGGAGGAGGCTCAAAGCGCTCCAT 731
Db 188 GlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAlaArgGluLeuLysAlaLeuHis 207
QY 732 CTGCTTCTCAGTACGATGACAGAGCAGTTTCCCTCAGTCAGAGCTCCCTTAAGCCAATG 791
Db 208 LeuLeuProGlnTyrAspAspGlnSerSerPheProGlnSerGluLeuProLysPromet 227
QY 792 ACAACTTTAGTGGGAGAGCTTCTGCCAGTACACGAGCAAGTAAATCTCATCACACAGGTT 851
Db 228 ThrThrLeuValGlyArgLeuLeuProValProAlaLysLeuAsnLeuIleThrGlnVal 247
QY 852 GATAATGGAGCTCTCCATCAGCTGTCATGAGGCGGCTGCTTCCCTCTTGACCTGCTCTG 911
Db 248 AspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPheProSerGlyProAlaLeu 267
QY 912 CAAGGGCCACCCAAATAACTCTGCTGGG 941
Db 268 GlnGlyProProLysIleThrLeuSerGly 277
RESULT 4
```



```
Db 45 AlaAlaProThrValValAlaThrSerValLysProIleLeu-----SerSerSerVal 62
Qy 765 CCTAGTCAGAGCTCCCTTAAGCCATGACAACTTACTAGGGAAGACATCTCCAGTACCA 824
Db 63 ProSerThrIleArgPro---GlyMetThrIleAlaIleGlyGlnValThrGlnValArg 81
Qy 825 GCGAAGTTAAATCTCATCACACAGGTGTGATAATGAGCTCTCCCATCAGCTGTCAATGGG 884
Db 82 ProThrLeuProMetAlaThrThrmSerAsnProProSerGlnSerGlnIleValAsn 101
Qy 885 GCT-----GCTTTCCCTCTGGAGCTCTCTGCAAGGCCA---CCCAATA 929
Db 102 AlaProIleArgHisProIleProGluSerProLysAlaArgGlyProArgProAsnVal 121
Qy 930 -----ACT 932
Db 122 GluGlyArgAspGlyThrProGlnLysLysLysLysCysAsnCysLysHisSerArgCys 141
Qy 933 CTGTCTGGGTACTGTGACTGCTTCTCCAGCGGGACTTCTGCAACAGCTGCAGCTGC--- 989
Db 142 LeuLysLeuTyrcysGluCysPheAlaSerGlyThrTyrcysAspGlyCysAsnCysVal 161
Qy 990 -----ACACACTCGCCATGAGCTCGAGCTCTCAAGCCATAAAGCGTGCTT 1040
Db 162 AsnCysPheAsnAsnValAspAsnGluProAlaArgArgGluAlaValGluAlaThrLeu 181
Qy 1041 GATGAAATCTGAAGCTTCCCAACCAAAATGGGAAA-----GGCGGT--- 1085
Db 182 GluArgAsnProPheAlaPheArgProLysIleAlaSerProHisGlyGlyArgAsp 201
Qy 1086 -----CTGGGAGCTGCTAAACTT-----CGACACAGCAAGGCTGCAACTGT 1127
Db 202 LysArgGluAspIleGlyGluValValLeuLeuGlyLysHisAsnLysGlyCysHisCys 221
Qy 1128 AAGCGCTCAGCGCTGCTGAAGAACTACTGTGAGTGCTATGAGGCAAAATCATGTGTTCT 1187
Db 222 LysLysSerGlyCysLeuLysTyrcysGluCysPheGlnAlaAsnIleLeuCysSer 241
Qy 1188 TCCATTTCCTCAATGCTGCTTCAAAACTATGAAGAACTCCAGCAACCAAAATGCTG 1247
Db 242 GluAsnCysLysCysLeuAspCysLysAsnPheGluGlySerGluGluArgGlnAlaLeu 261
Qy 1248 ATGAGCACACCCAC-----TACATGGAG----- 1271
Db 262 PheHisGlyGluHisSerAsnHisMetAlaTyLeuGlnGlnAlaAlaAsnAlaAlaIle 281
Qy 1272 CTGGGGACTTTGAGCAGCAGCCATTATTG---TCCCGAGCCAAAGTTCTCAGGACCTCCA 1328
Db 282 ThrGlyAlaValGlySerSerGlyPheAlaProSerProAla-----Pro 296
Qy 1329 AAACCTGAGAAAAATAGGCAG 1349
Db 297 LysArgArgLysGlyGlnGlu 303
RESULT 9
Q9CAV1
ID Q9CAV1 PRELIMINARY; PRT; 601 AA.
AC Q9CAV1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 66.8 kDa protein.
GN T9114.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
```

```
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Felsmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA Delsen M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Horischer K., Kaufer G., Loehner T.-H., Nordsiek G.,
RA Reichert J., Schaefer M., Schoen O., Bagues M., Terol J., Clement J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Coake R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarsee A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argirou A., Flores M., Licuori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Meyer K.F.X., Kaul S., Town C.D., Koo H.-L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pal G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida T.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RA thaliana.";
RL Nature 408:820-822(2000).
DR EMBL; AC009465; AAG51411.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
KW Hypothetical protein.
SQ SEQUENCE 601 AA; 56786 MW; 1084BE970D81E022 CRC64;
```

```
Alignment Scores:
Pred. No.: 1,21e-16 Length: 601
Score: 266.50 Matches: 89
Percent Similarity: 37.77% Conservative: 33
Best Local Similarity: 27.55% Mismatches: 96
Query Match: 6.65% Indels: 105
DB: Gaps: 16
```

US-09-743-237-1 (1-2241) x Q9CAV1 (1-601)

```
Qy 495 GTTCAATCAAGAGCAGGTGGTAGTGCCAGCGCGCCCTGAAGACGACGTTC 554
Db 236 ValGluGluArgGluGlySerCysSerVal-----GlnValAlaAla 249
Qy 555 CAGGCCCTCTGGCTCAGGAATCCTGT-----TGCAAGTTCCCATCATCCAG 602
Db 250 GlyAlaProAspIleAsnLeuSerCysSerSerLysValAlaAlaIleAspSerThra 269
Qy 603 GAGCAGGAGGAGGCTCCAGCTGCCCTCGGAAAGAAAGACTCCAGCCCATGGTATTGT 662
Db 270 GluAlaGluAspLysGluAsp-----LysAspLeuGlnPro----- 281
Qy 663 CAGCTGAAGAGGAGGCGCCAGACTCTGTGCATACACACTGTGGCGGAGGAGCTCAA 722
Db 282 -----SerGlyLysGlnArgSerValArgArgArgCysLeuThrPheAspMetGly 298
Qy 723 GCGCTCCAT---CTGCTTCCTCAGTACGATGACCCAGAGCAGTTTCCCTCAG----- 770
Db 299 GlySerHisLysArgIleProLeuArgAspSerThrAsnAspLeuProLeuAspSerThr 318
Qy 771 -----TCAGAGCTCCCTAAGCCCAATGACAACTTTA-----GTGGGA 806
Db 319 SerIleAsnLysAlaProSerProGlnAsnCysLeuAspThrSerLysGlnAspThrAsp 338
Qy 807 AGACTTCTGCCAGTACCAGCGCAAGTTAAATCTCATCACAGGTTGATAGGAGCTCTC 866
Db 339 GluIleLeuProIleProArgThrIleGlyLeuHisLeu-----AsnGlyPheVal 355
```


RN SEQUENCE FROM N.A.
 RX MEDLINE=98121113; PubMed=9461215;
 RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
 RA Bergkamp R., Dirkse W., van Stavoren M., Stiekema W., Drost L.,
 RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
 RA Wedler E., Wambutt R., Weitzenecker T., Pohl T.M., Terryn N.,
 RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
 RA Aubion K.D., Rieger M., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
 RA Antong S., Gy I., Kreis M., Schaeffer M., Funk B., Mueller-Auer S.,
 RA Slivey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
 RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
 RA Hilbert H., Duesterhoft A., Moeres T., Jones J.D.G., Eneva T.,
 RA Palme K., Benes V., Reichman S., Ansoorge W., Cooke R., Berger C.,
 RA Delsen M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
 RA Schueller C., Chalwatzis N.
 RA "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 RT Arabidopsis thaliana."
 RL Nature 391:485-488(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z97337; CAB10256.1; -;
 DR EMBL; AL161539; CAB78519.1; -;
 DR InterPro; IPR005172; CXC.
 DR Pfam; PF03638; CXC; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 658 AA; 72129 MW; 5BB99148E5E13A77 CRC64;
 Alignment Scores:
 Pred. No.: 6.84e-15 Length: 658
 Score: 248.50 Matches: 107
 Percent Similarity: 39.95% Conservative: 50
 Best Local Similarity: 27.23% Mismatches: 160
 Query Match: 6.20% Indels: 76
 DB: 10 Gaps: 18
 US-09-743-237-1 (1-2241) x 023333 (1-658)
 QY 441 GACTCGATGACTACTGCCA---GGTCCCAAGCCAGCAAGCAAGCAAGAAAGTGT 497
 Db 186 AspAlaSerGluLeuLeuIlePheAspSerProAspAlaSerGluAlaPheArgCysPhe 205
 QY 498 GAAATCAAGAAGAGGTGGTAGTGCCAGCGCCAGCTGAGAGCGAGCTTCCAG 557
 Db 206 MetMetGlnArgAlaSerAsnSer-----GluAlaArgPheArg 218
 QY 558 GCCCTCTGGTCCAGGAATCTGTTCAGTTCATCATCCAGGAGGAGGAGGCC 617
 Db 219 AsnGlyValGluGlyGlnThr---MetGlnHisAspSerAsnLysGluProGluSerAla 237
 QY 618 TCCAGTCCCTCCGCAAGAAAGACTCCAGCCCCATGGTGCATTGTTCAG----- 665
 Db 238 AsnAlaIleProTyrGluValAsnSerGly-----ValIleSerGlnAlaValSerLeu 255
 QY 666 CTGAAGAGGAGCGCCAGCATGCTCTGCATAGC----- 698
 Db 256 LeuHisArgGlyIleArgArgCysLeuAspPheGluMetProGlyAsnLysGlnThr 275
 QY 699 -----ACTGTGGCGCGAGGAGCTCAAGCG----- 725
 Db 276 SerSerGluAsnAsnThrAlaAlaCysGluSerSerSerArgCysValValProSerIle 295
 QY 726 ---CTCCATCTG-----CTTCCTCAGTACGATGACGAGCAGGATTCCTCCAGTCA 773
 Db 296 GlyLeuHisLeuAsnAlaIleLeuMetSerSerLysAspCysLysThrAsnValThrGln 315
 QY 774 GAGCTCCCTAAAGCAACATTTAGTGGGAAGACTTCTGCCAGTACCAGCGAAGTTA 833
 Db 316 AspTyrSerCysSerAlaAsnIleGlnValGlyLeuGlnArgSerIleSerThrLeuGln 335
 QY 834 AATCTCATCACACAGGTTGTAATGGAGCTCTCCATCAGCTGTCAATGGGGCTGCCTTT 893

Db 336 AspSerLeuAspGlnThrGluAsnGluIleArgGluAspAlaAspGlnAspVal----- 353
 QY 894 CCCTCTGGACCTGCTCTGCAA-----GGGCCACCCAAAATAACTGTGCT 938
 Db 354 ProValGluProAlaLeuGlnGluLeuAsnLeuSerSerProLysLys-----Lys 370
 QY 939 GGGTACTGTGACTGCTCTCCAGGGGAGCTTCGC---AACAGCTGCAGCTGC----- 989
 Db 371 SerTyrCysGluCysPheAlaAlaGlyValTyrCysIleGluProCysSerCysIleAsp 390
 QY 990 -----ACAACCTCGCGCATGAGCTCGAGCGTTTCAAGCCATAAAGCGCTGTTCAT 1043
 Db 391 CysPheAsnLysProIleHisGluAspValValLeuAlaThrArgLysGlnIleGluSer 410
 QY 1044 AGAATCTGAAGCTTCCCAACCAAAATGGGAAGGCCCTCTGGAGCT----- 1094
 Db 411 ArgAsnProLeuAlaPheAlaProLysValIleArgAsnSerAspSerValGlnGluThr 430
 QY 1095 -----GCTAAACTTCGACACAGCAAGGGTGCACACTGTAAG 1130
 Db 431 GlyAspAlaSerLysThrProAlaSerAlaArgHisLysArgGlyCysAsnCysLys 450
 QY 1131 CGTCAAGCTCCCTGAGAACTACTGTGAGTGTATGAGGCCAAATCATGTCTTCC 1190
 Db 451 LysSerAsnCysLeuLysLysTyrCysGluCysTyrGlnGlyValGlyCysSerIle 470
 QY 1191 ATTTGCAAAATGCTTGTTCAAAACCTATCAAGAAAGTCCAGCAAGCAAAATGCTGATG 1250
 Db 471 AsnCysArgCysGluGlyCysAsnAlaPheGlyArgLysAspGlySerSerIleAsp 490
 QY 1251 AGCACACCCCTACTACATGGAGCTGGGAGCTTGTGAGAGCAGCATTAATTTGCCCCAGCC 1310
 Db 491 MetGluAlaGluGlnGlu-----GluAsnGluThrSerGluLysSerArgThrAla 508
 QY 1311 AAGTCTCAGGACCTCCAAAACCTAGAAAATAAGCAGGAGCTTCT----- 1356
 Db 509 Lys-SerGln-----GlnAsnThrGluValLeuMetArgLysAspMetSerSerAlaLe 526
 QY 1357 -CCTGTATCTCCTGGGAAGTAGTGGAGCCACATGCTGCTGCTGCTGCCCGCAGGTGAG 1415
 Db 526 uProThrThrProThrProIleTyrArgProGluLeuValGlnLeu---PropheSerSe 545
 QY 1416 GAAGCAGCAGGAGCAGCTGTTCGCCAAGCTGGCTG 1452
 Db 545 rSerLysAsnArgMetProProGlnSerLeuLeu 557
 RESULT 12
 Q92S22 ID Q92S22 PRELIMINARY; PRT; 896 AA.
 AC Q92S22;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Cysteine-rich polycomb-like protein.
 GN CPPI.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_TaxID=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20345129; PubMed=10859345;
 RA Cvitanich C., Pallisgaard N., Nielsen K.A., Hansen A.C., Larsen K.,
 RA Phakaski-Maunsbach K., Marcker K.A., Jensen E.O.;
 RT "CPPI, a novel type DNA-binding protein involved in the expression of
 RL a soybean leghemoglobin c3 gene."
 DR EMBL; AJ010165; CAA09028.1; -;
 DR InterPro; IPR005172; CXC.
 DR Pfam; PF03638; CXC; 2
 SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;

```
Alignment Scores:
Pred. No.: 1,18e-14 Length: 896
Score: 246.50 Matches: 70
Percent Similarity: 38.66% Conservatives: 34
Best Local Similarity: 26.02% Mismatches: 93
Query Match: 6.15% Indels: 72
DB: 10 Gaps: 8

US-09-743-237-1 (1-2241) x Q94AI2 (1-896)
QY 554 CCAGGCCCTCTGGCTCAGGAATCTGTTGCAA-----GTTCCCATCATC 598
DB 364 ProLysProSerGlyIleGlyLeuHisLeuAasnSerIleLeuAasnAlaIleProIleAasp 383
QY 599 CCAGNGGAGAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCATGTGTAT 658
DB 384 GlnAlaAlaThrThrGlyValArgLeu----- 392
QY 659 TTGTCACTGAAAGAGGCGCCAGATGCTCTGCATAGACAACTGTGTGGCGGAGGAGCT 718
DB 393 -----SerAspSerSer 396
QY 719 CAAGCGCTCCATCTGCTTCTCAGTACGATGACCAAGACAGTTCCTCCAGTCAGAGCT 778
DB 397 GlnGlyMetLysSerThrSerSerIleArg---LeuGlnArgMetGluAasnValLysArg 415
QY 779 CCTAAGCAATGACAACTTTAGTGGGAAGACTTCTGCCAGTACCAAGCAAGTTAAATCT 838
DB 416 SerIleLeuSerSerAasnValAspGlyArgSerLeuValAspThrArg-----ThrGlu 433
QY 839 CATCACAGGTGTATAATGGAGCTCTCCATCAGCTGTGTAATGGGCTGCTTCCCTC 898
DB 434 SerHis-GluIleAaspThrValAlaThrAspThrGlyAasnSerGluAasnGln 453
QY 899 TGGACTGCTCTGCAAGGCGCCACCAATAACT----- 932
DB 453 nProProSerProCysLysLysLysLysLysThrSerValThrAlaAaspAsnGlyCys 473
QY 933 -----CTGTCTGGGTACTGTGACTGTGCTTCTCCAG 961
DB 473 sLysArgCysAasnCysLysSerLysCysLysLysLysLysLeuTyrcysAaspPheAlaAl 493
QY 962 CGGGGACTTCTGC---AACAGCTGCAGCTGCACAACTCGCGCCATGAGCTCGAGCGCTT 1018
DB 493 aglyThrTyrcysThrAspProCysAlaCysGlnGlyCysLeuAasnArgProGluTyVa 513
QY 1019 CAAGCCATA-----AAGCGGTCTTGTATAGAAATCTCGAAGCTTTCCAACC 1066
DB 513 lGluThrValValGluThrLysGlnGlnIleGluSerArgAasnProIleAlaPheAlaPr 533
QY 1067 AAAAAATG-----GGGAAAGCGCTGTGG 1090
DB 533 OlyIleValGlnProThrThrAspIleSerSerHisMetAaspAspGluAasnLeuThrTh 553
QY 1091 AGCTGCTAAACTTCGACACAGCAAGGGTGCACACTTAAGCGCTCAGGCTCGCTGAAGAA 1150
DB 553 rProSerSerAlaArgHisLysArgGlyCysAasnCysLysArgSerMetCysLeuLysLy 573
QY 1151 CTACTGTAGTCTATGAGCCCAAAATCATGTCTTCCATTTGCAATGCAATGCTTGTG 1210
DB 573 sTyrcysGluCysTyrglnAlaasnValGlyCysSerSerGlyCysArgCysGluGlyCys 593
QY 1211 CAAAACTATGAAGAAAGTCCAGAA 1235
DB 593 sLysasnValHisGlyLysLysGlu 601

RESULT 13
Q94AI2
ID Q94AI2 PRELIMINARY: PRT: 526 AA.
AC Q94AI2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DB 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
```

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Putative DNA binding protein.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Gibson H.A.,

RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,

RA Tang C., Toriumi M., Yu G., Yu S., Bowser L., Carninci P., Chen H.,

RA Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,

RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,

RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,

RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,

RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,

RT "Full Length cDNA of gene MWI23.15/AT3g22780 (GI:9279696).";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AV046019; AAK76693.1;

DR InterPro: IPR005172; CXC.

DR Pfam: PF03638; CXC; 2.

SQ SEQUENCE 526 AA; 57886 MW; C6CD9684D417A9D9 CRC64;

Alignment Scores:

Pred. No.: 6.15e-13 Length: 526

Score: 228.00 Matches: 95

Percent Similarity: 36.64% Conservatives: 49

Best Local Similarity: 24.17% Mismatches: 137

Query Match: 5.69% Indels: 112

DB: 10 Gaps: 18

US-09-743-237-1 (1-2241) x Q94AI2 (1-526)

QY 366 AACTAGGAGAACCTCTGATTCCTCTCTACATCTCTAGTGTAGGTACAGGGG 425

DB 77 AsnGluAasnGlnProLeuAlaValLeuPro----- 86

QY 426 AATCATTTTTGGTGAGACTCCGATGAACTACTGCCAGGTTCCTCCAGGCAAGCA 485

DB 87 -----ThrAsnGluSerValPheAsnLeuHisArgGlyMetArg 100

QY 486 AGAAAAAGTGTGAAATCAAGAACGAGGTGGTGTGTGCCAGCGCCAGCCCTGAAGAC 545

DB 101 ArgArgCysLeuAaspPheGlu-----MetProGlyLysArgLysLysAsp 115

QY 546 GCAGCTTTCAGGCCCTCTGGCTCAGGAATCTGTGTGCAAGTTCCTCCATCCTCCAGGAG 605

DB 116 IleValAspAsp-----GlnGlnSerValCys-----AspAsnAsnValAla 129

QY 606 GCAGAGGAGCGCTCCAGCTGC----- 626

DB 130 GlyGluSerSerSerSerCysValValProGlyIleGlyLeuHisLeuAasnAlaValAla 149

QY 627 CCTCGAAGAAAGACTCCAGCCCATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 696

DB 150 MetSerAlaLysaspSerAsnIleSerValIleHisGlyTyrcysSerIleSerGlyGluIle 169

QY 687 CTCTCATAGACAACTGTGTGGCGGAGGAGCTCAAGCGCTCCTCATCTGCTTCTCTAGTAC 746

DB 170 GlnLysSerPheSerGlySerThrThrProIleGlnSerGlnAspThrValGlnGluThr 189

QY 747 GATGACCAAGC---AGTTTCTCTCAGTCAGAGCTCCCTAAGCCATGACAACTTGTAGTC 803

DB 190 SerAspGlnAlaGluAasnGluProValGluGluValProLysAlaLeu----- 205

QY 804 GGAAGACTTCTGCCAGTACCAAGTAAATCTCATCACAGCTTGTATATGAGCT 863

DB 206 -----MetPhePro-----GluLeuAasnLeu-----GlySer 214

QY 864 CTCCCATCAGCTGTCAATGGGCTGCTTTCCTCTGGACCTCTCTGCAAGG----- 917

DB 215 LeuLysLysLysMetArg-----LysSerGluGlnAlaGlyGluGlyGluSer 230


```

QY 918 -----CCACCAAAATAACTGTCTGGGTACTGTGACTGCTTCTCC 959
Db 231 CysLysArgCysAsnCysLysLysSerLysCysLeuLysLeuTyrCysGluCysPheAla 250
QY 960 AGCGGGGACTTCTGC---AACAGCTGCAGCTGC-----AACACCTCGCCCAT 1004
Db 251 AlaGlyValTyrCysIleGluProCysSerCysIleAspCysPheAsnLysProIleHis 270
QY 1005 GAGCTCGAGCCCTCAAGGCCATAAAGCGTGTCTTGATAGAAATCTGAAGCTTTCCAA 1064
Db 271 GluGluThrValLeuAlaThrArgLysGlnIleGluSerArgAsnProLeuAlaPheAla 290
QY 1065 CCAAAATG-----GGGAAAGGCCGTCTGGGA 1091
Db 291 ProLysValIleArgAsnAlaAspSerIleMetGluAlaSerAspAlaSerLysThr 310
QY 1092 GCTGCTAACTTCGACAGACGAAGGTGCAACTGTGAAGCGCTCAGGCTGCCTCAAGAAC 1151
Db 311 ProAlaSerAlaArgHisLysArgGlyCysAsnCysLysLysSerAsnCysMetLysLys 330
QY 1152 TACTGTGAGTCTATGAGGCCAAATCATGTGTTCTTCCATTTCGAATGCCATTGCTTGC 1211
Db 331 TyrCysGluCysTyrGlnGlyGlyValGlyCysSerMetAsnCysArgCysGluGlyCys 350
QY 1212 AAAAAC-----TATGAA 1223
Db 351 ThrAsnValPheGlyArgLysAspGlySerLeuLeuValIleMetGluSerLysLeuGlu 370
QY 1224 GAAGTCCAGAA-----CGAAATACTGTATGACACACCCCTACATGAG- 1271
Db 371 GluAsnGlnGluThrTyrGluLysArgIleAlaLysIleGlnHisAsnValGluValSer 390
QY 1272 -----CTGGGGACTTTTCAGAGCAGCCATTATTCTCCCGACGCCAAG 1313
Db 391 LysGluValGluGlnAsnProSerSerAspGlnProSerThrProLeuProProTyrArg 410
QY 1314 -----TTCTCAGGACCTCCAAACTCGAGAAATAAGG 1346
Db 411 HisLeuValValHisGlnProPheLeuSerLysAsnArg 423

RESULT 14
Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CXC domain protein TS01 (Putative DNA binding protein).
GN TS01.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LER;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TS01 is a novel protein that modulates cytokinesis and cell expansion
RL in Arabidopsis.";
RL Development 127:2219-2226(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song Y.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TS01, an
RL Arabidopsis gene with cysteine-rich repeats.";
RL Development 127:2207-2217(2000).
DR EMBL; AF204059; AAF69124.1;
DR EMBL; AF206324; AAF27433.1;

```

```

DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA8550F6E5D4 CRC64;

Alignment Scores:
Pred. No.: 6,71e-13 Length: 695
Score: 228.00 Matches: 91
Percent Similarity: 35.58% Conservative: 46
Best Local Similarity: 23.64% Mismatches: 152
Query Match: 5.69% Indels: 96
DB: 10 Gaps: 14

US-09-743-237-1 (1-2241) x Q9LE32 (1-695)
QY 366 AACTTAGAGAACTCTGATTCTCTCGCTCTACATCCTGAGTCTAGGTGTACAGGGG 425
Db 246 AsnGluAsnGlnProLeuAlaValLeuPro----- 255
QY 426 AAATCATTTTGGTGAGACTCCGATGAATCTGCCAGGTTCCCAAGGCGCAGCAAGCA 485
Db 256 -----ThrAsnGluSerValPheAsnLeuHisArgGlyGlyMetArg 269
QY 486 AGAAAAAGTGTGAATAACAAGAGCAGGTGTAGTGTCCAGCGCGCCCTCGAAGAC 545
Db 270 ArgArgCysLeuAspPheGlu-----MetProGlyLysArgLysLysAsp 284
QY 546 GCAGCTTTCCAGGCCCTCTGGCTCAGGAATCTGTTCAGAGTTCCCATCATCCAGGAG 605
Db 285 IleValAspAsp-----GlnGlnSerValCys-----AspAsnValAla 298
QY 606 GCAGAGGAGGCTCCAGCTGC----- 626
Db 299 GlyGluSerSerSerCysValValProGlyIleGlyLeuHisLeuAsnAlaValAla 318
QY 627 CCTCGGAAGAAAGACTCCAGCCCATGGTGTATTTGTAGCTGAAGAGGCGCCAGATG 686
Db 319 MetSerAlaLysAspSerAsnIleSerValIleHisGlyTyrSerIleSerGlyGlu 338
QY 687 CTCTGCATACAACTGTGGCGGAGGAGCTCAAAGCGCTCCATCTGCTTCTCCTCAGTAC 746
Db 339 GlnLysSerPheSerGlySerThrThrProIleGlnSerGlnAspThrValGlnGluThr 358
QY 747 GATGACGAGAGC---AGTTTCCCTCAGTCAAGCTCCCTTAAGCCAATGACAACTTTAGT 803
Db 359 SerAspGlnAlaGluAsnGluProValGluGluValProLysAlaLeuValPheProGlu 378
QY 804 GGAAGACTTCTGCCAGTACCAGCAAGTAAATCTCATCACACAGGTGATAAATGGAGCT 863
Db 379 LeuAsnLeuGlySerLeuLysLysLysMetArgLysSerGluGlnAlaGlyGluGlyGlu 398
QY 864 CTCCCATCAGCTGTCAATGGGGCTGCCTTTCCCTCTGGACCTGTCTGCAAGGCCACCC 923
Db 399 SerCysLysArgCysAsn-----CysLysLys 407
QY 924 AAATAACTCTGTGGGTACTGTGACTGTCTTCCAGCGGGGACTTCTGC---AACAGC 980
Db 408 SerLysCysLeuLysLeuTyrCysGluCysPheAlaAlaGlyValTyrCysIleGluPro 427
QY 981 TGCAGCTGC-----AACCACTGGCCATGAGCTCGAGCGCTTCAAGGCCATA 1028
Db 428 CysSerCysIleAspCysPheAsnLysProIleHisGluGluThrValLeuAlaThrArg 447
QY 1029 AAGCGGTCTTGTATAGAAATCTCGAAGCTTTCCAAACCAAAATG----- 1073
Db 448 LysGlnIleGluSerArgAsnProLeuAlaPheAlaProLysValIleArgAsnAlaAsp 467
QY 1074 -----GGGAAAGGCCGTCTGGAGCTGCTAACTTCGACACAGCAAA 1115
Db 468 SerIleMetGluAlaSerAspAspAlaSerLysThrProAlaSerAlaArgHisLysArg 487
QY 1116 GGGTGCACACTGTAAAGCGCTCAGGCTGCTGAAGAACTACTGTGAGTGTATGAGTATGAGCCCAAA 1175
Db 488 GlyCysAsnCysLysLysSerAsnCysMetLysLysTyrCysGluCysTyrGlnGlyGly 507

```



```

QY 1176 ATCATGTGTTCTTCCATTGCAAAATGCTGCTTGCAGAAAC----- 1217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 ValGlyCysSerMetAsnCysArgCysGluGlyCysThrAsnValPheGlyArgLysAsp 527
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1218 -----TATGAAGAAAGTCCAGAA-----CGAAAA 1241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 GlySerLeuValIleMetGluSerLysLeuGluGluAsnGlnGluThrTyrGluLys 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1242 ATGCTGATGAGACACACCCCACTACATGGAG-----CCTGGG 1277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 ArgIleAlaLysIleGlnHisValGluValSerLysGluValGluGlnAsnProSer 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1278 GACTTTGAGACGACCATTTATTTGCCCCAGCAAG-----TTCTCAGGACCTCCAAA 1331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 568 SerAspGlnProSerThrProLeuProTyrArgHisLeuValValHisGlnProPhe 587
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1332 CTGAGAAAAATAGG 1346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 LeuSerLysAsnArg 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
Q9LUI3
ID Q9LUI3 PRELIMINARY; PRT; 695 AA.
AC Q9LUI3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE DNA binding protein-like.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
DR EMBL; AB022223; BAB01253.1; -.
DR InterPro; IPR005172; CXC.
DR Pfam; PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76276 MW; 34BBA0E450F6BCE1 CRC64;

Alignment Scores:
Pred. No.: 6,71e-13 Length: 695
Score: 228.00 Matches: 91
Percent Similarity: 35.58% Conservative: 46
Best Local Similarity: 23.64% Mismatches: 152
Query Match: 5.69% Indels: 96
DB: 10 Gaps: 14

US-09-743-237-1 (1-2241) x Q9LUI3 (1-695)
QY 366 ACTTAGGAGAACCTCTGATTCCTGCTCATCTACATCTGCTAGTGTACAGGGGG 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 246 AsnGluAsnGlnProLeuAlaValLeuPro----- 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 426 AATCATTTTGTGTGAGACTCCCATGACTGCGAGGTCCAGGTTCCCAAGGCAAGCAAGCA 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 -----ThrAsnGluSerValPheAsnLeuHisArgGlyGlyMetArg 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 AGAAAAAGTTGAATCAAGAAGCAGGTGTAGTGTGCCAGGCGGCCTCGAAGAC 545
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: April 21, 2003, 11:38:24
Job time : 162.77 secs

```

Db 270 ArgArgCysLeuAspPheGlu-----MetProGlyLysArgLysLysAsp 284
QY 546 GCAGCTTTCCAGCGCCCTCTGCTCAGGAATCCTGTTGCAAGTCCATCATCCAGGAG 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 IleValAspAsp-----GlnGlnSerValCys-----AspAsnAsnAla 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 606 GCAGGAGGAGCCTCCAGCTGC----- 626
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 GlyGluSerSerSerCysValValProGlyIleGlyLeuHisLeuAsnAlaValAla 318
QY 627 COTCGAAGAAAGACTCCAGCCCATGTTGTTGCTAGCTGAAAGAGGCGCCAGATG 686
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 MetSerAlaLysAspSerAsnIleSerValIleHisGlyTyr-SerIleSerGlyGluIle 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 687 CTCTCATAGACAACCTGTGGCGGAGGAGCTCAAGAGCCTCATCTGCTCTCAGTAC 746
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 GlnLysSerPheSerGlySerThrThrProIleGlnSerGlnAspThrValGlnGluThr 358
QY 747 GATGACCGAGC---AGTTTCCCTCAGTCAGAGCTCCCTAAGCCMATGACAACCTTTAGTG 803
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 SerAspGlnAlaGluAsnGluProValGluGluValProLysAlaLeuValPheProGlu 378
QY 804 GGAAGACTTCTCCAGTACCAGCAAGTTAAATCTCATCACACAGCTTGATATGGAGCT 863
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 LeuAsnLeuGlySerLeuLysLysMetArgLysSerGluGlnAlaGlyGluGlyGlu 398
QY 864 CTCCTCATCAGCTGTCATGGGGCTGCTTCCCTCTGGACCTCTCTGCAAGGGCCACCC 923
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 SerCysLysArgCysAsn-----CysLysLys 407
QY 924 AAAATAACTCTCTGCTGGTACTGTGACTGCTTCCAGCGGGGACTTCTGC---AACAGC 980
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 SerLysCysLeuLysLeuTyrCysGluCysPheAlaIleGlyValTyrCysIleGluPro 427
QY 981 TCAGCTGC-----AACAACTGCGCCATGAGCTCAGCGCTTCAAGGCCATA 1028
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 CysSerCysIleAspCysPheAsnLysProIleHisGluGluThrValLeuAlaThrArg 447
QY 1029 AAGGGTGTCTTGATAGAAATCCTGAGCTTTCCCAACCAAAAATG----- 1073
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 448 LysGlnIleGluSerArgAsnProLeuAlaPheAlaProLysValIleArgAsnAlaAsp 467
QY 1074 -----GGGAAAGGCCCTCTGGGAGCTGCTAAACTTCGACACAGCAAA 1115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 SerIleMetGluAlaSerAspAlaSerLysThrProAlaSerAlaArgHisLysArg 487
QY 1116 GGTCGCAACTGAAGCGCTCAGGCTGCCTGAAGAACTACTGTGAGTGTATGAGGCCAAA 1175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 GlyCysAsnCysLysLysSerAsnCysMetLysLysTyrCysGluCysTyrGlnGlyGly 507
QY 1176 ATCATGTGTTCTTCCATTGCAAAATGCAATGCTTGCAGAAAC----- 1217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 ValGlyCysSerMetAsnCysArgCysGluGlyCysThrAsnValPheGlyArgLysAsp 527
QY 1218 -----TATGAAGAAAGTCCAGAA-----CGAAAA 1241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 GlySerLeuValIleMetGluSerLysLeuGluGluAsnGlnGluThrTyrGluLys 547
QY 1242 ATGCTGATGAGCACACCCCACTACATGGAG-----CCTGGG 1277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 ArgIleAlaLysIleGlnHisAsnValGluValSerLysGluValGluGlnAsnProSer 567
QY 1278 GACTTTGAGAGAGCATTATTTGCCCCAGCAAG-----TTCTCAGGACCTCCAAA 1331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 568 SerAspGlnProSerThrProLeuProTyrArgHisLeuValValHisGlnProPhe 587
QY 1332 CTGAGAAAAATAGG 1346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 LeuSerLysAsnArg 592
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:49:14 : Search time 92.5 Seconds
(without alignments)
6148.256 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 3824
Sequence: 1 aattcgggggtcaaggcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09743237/runat_21042003_112208_1454/app_query.fasta_1.2311
-DB=A_Geneseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237.eccn_1_153@runat_21042003_112208_1454 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGOQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	1647	43.1	299	21	AA168464	Human testis speci
2	1251	32.7	295	21	AA168463	Mouse testis speci
3	836.5	21.9	251	22	AA168462	Human polypeptide
4	730	19.1	147	22	AA168461	Human testicular a
5	730	19.1	147	22	AA168460	Human reproductive
6	483.5	12.6	950	22	AA168459	Drosophila melanog
7	441	11.5	438	20	AA168458	Caenorhabditis ele
8	378.5	9.9	280	22	AA168457	Novel human diagno
9	278	7.3	53	22	AA168456	Human colon cancer
10	200	5.2	243	22	AA168455	Drosophila melanog
11	149	3.9	4561	22	AA168454	Novel human diagno
12	149	3.9	4561	22	AA168453	Novel human diagno
13	130.5	3.4	2743	23	AA168452	Human laminin alph
14	130.5	3.4	3695	23	AA168451	Human laminin alph
15	130.5	3.4	3695	23	AA168450	Human laminin alph
16	130.5	3.4	3705	23	AA168449	Human laminin alph
17	125	3.3	1037	22	AA168448	Mouse CRIM1 protei
18	124	3.2	3011	13	AA168447	Compiled HCV sequ
19	121	3.2	507	16	AA168446	HCV J1 NS3-NS4 dom
20	121	3.2	1017	22	AA168445	TuT protein #4
21	120	3.2	1615	22	AA168444	Protein #3 encoded
22	120	3.1	3010	14	AA168443	Polypeptide coded
23	120	3.1	3011	14	AA168442	Hepatitis C virus
24	119.5	3.1	1035	22	AA168441	Drosophila melanog
25	119.5	3.1	3010	16	AA168440	Partial HCV non-st
26	119.5	3.1	3010	16	AA168439	Hepatitis C virus
27	119	3.1	1036	23	AA168438	Human pharmaceutical
28	118.5	3.1	1801	19	AA168437	Rat laminin B2 cha
29	118	3.1	732	22	AA168436	Human NOV10 protei
30	118	3.1	2515	22	AA168435	Drosophila melanog
31	117	3.1	1021	18	AA168434	HCV antigen combin
32	117	3.1	1021	19	AA168433	Fusion protein c20
33	117	3.1	1021	23	AA168432	PSOD/c200/core exp
34	117	3.1	1771	22	AA168431	HCV NS35 polypepti
35	117	3.1	1771	22	AA168430	HCV NS35 polypepti
36	117	3.1	1771	22	AA168429	Amino acid sequenc
37	117	3.1	1771	22	AA168428	Amino acid sequenc
38	117	3.1	1892	22	AA168427	Amino acid sequenc
39	117	3.1	1911	22	AA168426	Amino acid sequenc
40	117	3.1	1921	22	AA168425	Amino acid sequenc
41	117	3.1	1944	22	AA168424	Peptide encoded by
42	117	3.1	2261	10	AA168423	Sequence encoded by
43	117	3.1	2301	10	AA168422	HCV polypeptide 1
44	117	3.1	2435	13	AA168421	Sequence encoded 1
45	117	3.1	2436	10	AA168420	Sequence encoded 1

ALIGNMENTS

RESULT 1
AA168464
ID AA168464 standard; Protein; 299 AA.
XX AA168464;
AC AA168464;
XX
DT 25-APR-2000 (first entry)
XX
DE Human testis specific factor tesmin SEQ ID NO:5.
XX
KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.
XX
OS Homo sapiens.
XX
PN WO200004147-A1.
XX
PD 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-JP03859.
 XX PR 17-JUL-1998; 98JP-0219856.
 XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX DR WPI: 2000-147785/13.
 XX DR N-PSDB: AAZ88156.
 XX PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 XX PT useful for investigation of germ cell differentiation and sterility -
 XX PS Claim 1; Page 50-52; 63pp; Japanese.
 XX CC The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX SQ Sequence 299 AA;
 Alignment Scores:
 Pred. No.: 6.48e-151 Length: 299
 Score: 1647.00 Matches: 299
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.07% Indels: 0
 DB: 21 Gaps: 0
 US-09-743-237-3 (1-2134) x AAY68464 (1-299)
 QY 407 ATGTGATATGCCAATTAAGAGGGGACACAAATGCTATGTATAGACAATTCTAGACA 466
 DB 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
 QY 467 AGAAGCTAAAGCACTCCATTGGTTCCTCAGTATCAAGATCAAAATAATTATCTACAG 526
 DB 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnTyrLeuGln 40
 QY 527 TCAGATGTCCTAAACCAATGACTGCTTTAGTAGGAGATTTTCCAGCATCAACAAA 586
 DB 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
 QY 587 TTAATCTCATACACAACACTTGAGGAGCCCTTACATCGGTAGTCAACGGTCTGCT 646
 DB 61 LeuAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAla 80
 QY 647 TTCCCTCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGTACTGTGAC 706
 DB 81 PheProSerGlySerThrLeuProGlyProProLysIleThrLeuAlaGlyTyrCysAsp 100
 QY 707 TGCTTTGCCAGTGGGACTTTTGCACAACTGCAATGTAATAATTGTGCAACAACCTG 766
 DB 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnCysAsnAsnCysAsnLeu 120
 QY 767 CATCATGATATTGAACGGTTTAAGCCATTAAGCATGCTTTGTTAGAAATCCAGAGCT 826
 DB 121 HisHisAspIleGluArgPheLysAlaIleLysAlaCysLeuGlyArgAsnProGluAla 140
 QY 827 TTCAGCCAAAATTTGGAGGGCCCAATGGCCAATGTCAGCCCGCCAGCACAAAGGG 886
 DB 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
 QY 887 TGCAACTGCAGGAGTCCAGTCCGCTGAAGAATTAAGTCCAGTCTATGAGGCCCAAT 946
 DB 161 CysAsnCysArgArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaGlnIle 180
 QY 947 ATGTGTTCTTCTATTTCGAATGCTATGGTTGCAAAATTAATTAAGAAAGCCCAACGA 1006
 DB: 21 Gaps: 0

DB 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200
 QY 1007 AAGACACTAATGACATGCCAACTACATGACAGCTGAGCTTTGGAGGACGACCATTAC 1066
 DB 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
 QY 1067 CTGCCACCAACGAAATTTTCAGGACTTCCAAAGATTCAGTCACGATAGCGGCTTCCTCA 1126
 DB 221 LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgArgProSerSer 240
 QY 1127 TGCATCTCTGGGAGGTGGAGGCCACATGCCCTGCTGCTCAGGGAGAAGAG 1186
 DB 241 CysIleSerTrpGluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 260
 QY 1187 GCCGAGAAAGAACACTCTCCAACTGCTGCAGCAGCAGATGATCTCGGAGGAATTTGGA 1246
 DB 261 AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluGluPheGly 280
 QY 1247 AGTGTCTTATCACAGATTCCTCCACACTGAGTTTAAATTAAGGATTGAAAATGGAG 1303
 DB 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299
 RESULT 2
 AAY68463
 ID AAY68463 standard; Protein; 295 AA.
 XX AC AAY68463;
 XX DT 25-APR-2000 (first entry)
 XX DE Mouse testis specific factor tesmin SEQ ID NO:4.
 XX KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 XX KW differentiation regulatory factor; male germ cell regulatory actor;
 XX KW germ cell differentiation; sterility.
 XX OS Mus musculus.
 XX PN WO200004147-A1.
 XX PD 27-JAN-2000.
 XX PF 16-JUL-1999; 99WO-JP03859.
 XX PR 17-JUL-1998; 98JP-0219856.
 XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX DR WPI: 2000-147785/13.
 XX DR N-PSDB: AAZ88155, AAZ88157.
 XX PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 XX PT useful for investigation of germ cell differentiation and sterility -
 XX PS Claim 1; Page 47-49; 63pp; Japanese.
 XX CC The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX SQ Sequence 295 AA;
 Alignment Scores:
 Pred. No.: 1.92e-112 Length: 295
 Score: 1251.00 Matches: 228
 Percent Similarity: 84.95% Conservative: 26
 Best Local Similarity: 76.25% Mismatches: 41
 Query Match: 32.71% Indels: 4
 DB: 21 Gaps: 1

US-09-743-237-3 (1-2134) x AAY68463 (1-295)

```
QY 407 ATGGTATATGCAATTTGAAGGGGCGACACAAATCTATGTATAGACAAATTTAGAAC 466
Db 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 467 AGAGAACTAAAGCACTCCATTGGTCTCTCAGTATCAAGATCAAAATTAATTTATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuLeuProGlnTyrAspAsnGlnSerPheProGln 40
QY 527 TCAGATGTCCTTAACCAATGACTGCTTAGTAGGAGATTTTGGCCAGCATCAACAAA 586
Db 41 SerGluLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 587 TTAATCTCTATTACAAACAACTTGGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 647 TTCCCTCGGATCAACTCTTCCAGGACCAACAAATAAATTTGGTGGGTACTGTGAC 706
Db 81 PheProSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGlyTyrCysAsp 100
QY 707 TGTCTTGGCAGTGGGACTTTTGCAACAACATGCAATTTGTAATAATTTGTTCAACAACCTTG 766
Db 101 CysPheSerSerGlyAspPheCysAsnSerCysSer-----CysAsnAsnLeu 116
QY 767 CATCATGATATCAACGGTTTAAGCCATTAAGGCATGCTTGGTGAAGAATCCAGAGCT 826
Db 117 ArgHisGluLeuGluArgPheLysAlaIleLysAlaCysLeuAspArgAsnProGluAla 136
QY 827 TTCACGCCAAAATTTGGGAGGCCCAATTTGGCAATGTCAAGCCCGACACACAAAGGG 886
Db 137 PheGlnProLysMetGlyGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGly 156
QY 887 TCGAATCGAGGAGGTGAGCTGCTCGAAGAATTTACTGCGAGTGTATGAGGCCCAAT 946
Db 157 CysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIle 176
QY 947 ATGTGTTCTTCTATTGCAAAATGCATTTGTCAAAATATGCAAGAAATGCAAGCCAGAACGA 1006
Db 177 MetCysSerSerIleCysLysCysIleAlaCysLysAsnTyrGluLeuSerProGluArg 196
QY 1007 AGACACTAATGAGTGCACAACTACATGCAGACTGGAGGTTTGAAGGAGCCCATAC 1066
Db 197 LysMetLeuMetSerThrProHisTyrMetGluProGlyAspPheGluSerHisTyr 216
QY 1067 CTCCACCAACGAATTTTCAGGACTTCCAGATTCAGTACAGTACAGGCGGCTTCCCTCA 1126
Db 217 LeuSerProAlaLysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSer 236
QY 1127 TGCATCTCTGGGAGTGTGGAGGCCACATGCGCTGCTGCTGCTCAGGGAGGAAGAG 1186
Db 237 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 256
QY 1187 GCGGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGACAGATGATCTCGGAGGAATTTGA 1246
Db 257 AlaGluGlnGluHisCysSerProSerLeuAlaGluGlnMetIleLeuGluPheGly 276
QY 1247 AGTGCTTTATCACAGATTTCCACACTGAGTTTAATCTAAGGGATGAAATGGAG 1303
Db 277 ArgCysLeuSerGlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295
```

RESULT 3

ID AAM93348 standard; Protein; 251 AA.

XX AC AAM93348;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide, SEQ ID NO: 2895.

```
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
OS Homo sapiens.
PN EP1130094-A2.
XX 05-SEP-2001.
XX 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI: 2001-524255/58.
DR N-PSDB; AAK94268.
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation.
XX Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX Sequence 251 AA;
SQ
```

Alignment Scores:

Pred. No.:	3,36e-72	Length:	251
Score:	836.50	Matches:	167
Percent Similarity:	95.43%	Conservative:	0
Best Local Similarity:	95.43%	Mismatches:	5
Query Match:	21.88%	Indels:	3
DB:	22	Gaps:	1

US-09-743-237-3 (1-2134) x AAM93348 (1-251)

```
QY 9 GTCAAGCGNAGCTCCGGGGGCGACAGCGGGGAGGCTCTCGGGGAGTACCCC 68
Db 79 ValIysAlaLysLeuAlaGlyGlyAspSerAspGlyGlyLeuLeuGlyIlePro 98
QY 69 GGATCCCGACAGCTCAGCGGCTGGAGGAGCTCGCGTCTCGAGGCCCGCGAGCCGCC 128
Db 99 GlyIleProGluLeuSerAlaLeuGluAspValAlaLeuLeuGlnAlaProGlnPro 118.
QY 129 GCTGCAAGCTGCACTTCTCTGCTCTCTGCTGCTACCCGCGCAGCCGCGGGTGT 188
Db 119 AlaCysAsnValHisPheLeuSerSerLeuLeuProAlaHisArgSerProAlaValLeu 138
QY 189 TGCCCTCGGGCGGCTGGTCTGCGAAGGAGCTCCACCCGGGGCGTCCGCATGATCCC 248
Db 139 ---ProLeuGlyAlaTrpValLeu-GluGlyAlaSerHisProGlyValArgMetIlePr 157
QY 249 AGTTGAAATCAAGTGAAGTGGTACTACTACAGTAAATAATCCGGAAGAAGCAACTT 308
Db 157 oValGluIleLysGlu-AlaGlyGlyThrThrThrSerAsnAsnProGluGluAlaThrL 177
QY 309 TGCAGAAATCTTCTTGCTCAGGAATCTCTGTTGCAAGTTCCCATGTCCTCCAGGAGG 368
```

```
Db 177 euGlnAsnLeuLeuAlaGlnGluSerCysLysPheProSerSerGlnGluLeuGluA 197
|||||
Oy 369 ATGCCTCTCTGCTGTTCTTAAGAAAGATTCCAAACCCCAATGGTGATATGCCAATTGAAG 428
|||||
Db 197 spAlaSerCysSerLeuLeuLysLysAspSerAsnPrometValIleCysGlnLeuLysG 217
|||||
Oy 429 GGGGACACAAATGCTATGATAGACAATCTAGAACAAAGAGAACTAAAGCACTCCCAAT 488
|||||
Db 217 lyGlyThrGlnMetLeuLysLysAspAsnSerArgThrArgGluLeuLysAlaLeuHisL 237
|||||
Oy 489 TGGTCTCTAGTATCAACATCAAAATAATATCTACACTCA 529
|||||
Db 237 euValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSer 250
|||||
RESULT 4
ABB96025
ID ABB96025 standard; Protein; 147 AA.
XX
AC ABB96025;
XX
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen SEQ ID NO: 1409.
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic.
XX
OS Homo sapiens.
XX
XX WO200155317-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231988.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
```

```
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
XX useful for preventing, diagnosing and/or treating testicular cancer
XX
XX Claim 11; SEQ ID NO 1409; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
XX human testicular antigens, and fragments of their genomic sequences. The
XX sequences can be used in the treatment of cardiovascular, urinary system,
XX reproductive system, immune, respiratory, neurological and
XX gastrointestinal disorders, infections, and particularly cancer,
XX especially testicular cancers. The present sequence is a protein of the
XX invention.
XX
XX SQ Sequence 147 AA;
XX
XX Alignment Scores:
XX Pred. No.: 5,938-62 Length: 147
XX Score: 730.00 Matches: 144
XX Percent Similarity: 97.30% Conservative: 0
XX Best Local Similarity: 97.30% Mismatches: 3
XX Query Match: 19.09% Indels: 1
XX DB: 22 Gaps: 0
XX
XX US-09-743-237-3 (1-2134) x ABB96025 (1-147)
XX
XX QY 229 CCGGGCGTCGCGATGATCCAGTTGAAATCAAGGTAAGCAGGTGGTACTACAGTAA 288
XX 1 ProGlyValArgMetIleProValGluIleLysGlu-AlaGlyGlyThrThrSerAs 20
XX
XX QY 289 TAATCGGGAAGCAACATTTCTGTCAGGAATCCTGTTCAGGAATCCTGTTCAGGTTCC 348
XX 20 nAsnProGluGluAlaThrLeuGlnAsnLeuLeuAlaGlnGluSerCysCysLysPhePr 40
XX
XX QY 349 ATGGTCCCAAGCACTAGAGATGCGCTCTGCTCTCTTAGAAGATTCACCAACCAAT 408
XX 40 oSerSerGlnGluLeuGluAspAlaSerCysCysSerLeuLysLysAspSerAsnProMe 60
XX
XX QY 409 GGTGATATCCCAATTGAAGGGGGGCACAAATGCTATGATATAGACAATTCAGAACAG 468
```

```
Db 60 tValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThrAr 80
QY 469 AGAACTAAAGCACTCCATTGGTTCTCAGTATCAAGATCAAAATAATATCTACAGTC 528
Db 80 gGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnAsnTyrLeuGlnSe 100
QY 529 AGATGTCCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTCCCGAGCATCAACAAAT 588
Db 100 rAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLysLe 120
QY 589 AAATCTCATTACACAACTTGGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
Db 120 uAsnLeuIleThrGlnGlnLeuGluGlyAlaLeuProSerValValAsnGlySerAlaPh 140
QY 649 CCCCTCGGATCAACTCTTCCA 670
Db 140 eProSerGlySerThr***Pro 147
XX
XX RESULT 5
XX AAM95330
XX ID AAM95330 standard; Protein; 147 AA.
XX AC AAM95330;
XX XX
XX DT 21-NOV-2001 (first entry)
XX XX
XX DE Human reproductive system related antigen SEQ ID NO: 3988.
XX KW Human; reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy.
XX OS Homo sapiens.
XX PN WO200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
```



```
Db 668 nLeuGlnAlaGlnAlaLysGlnArgIleArgGlnGlnLeuProThrGluGlnSerTh 688
Qy 611 -----GAGGAGCCTTACCATCGGTA-----GTCAACGGGTC 642
Db 688 rProIleLysValGluProLysLeuProThrLeuProProGlyValLysAlaAsnValPr 708
Qy 643 TGCTTTCCCTCGGATCACTCTCCAGGACCAACCAATACTTTGGCTGGG----- 697
Db 708 oAlaLysProLeuPheGluValLeuLysProProAlaThrAlaAlaAlaGlyAlaVa 728
Qy 697 ----- 697
Db 728 lasProLeuGlyGlyMetThrSerArgLysHisCysAsnCysSerLysSerGlnCy 748
Qy 698 -----TACTGTGACTGCTTTGCCAGTGGGACTTTTGCACACACTGCAATTGTAA 747
Db 748 sLeuLysLeuTyrcysAspCysPheAlaAsnGlyGluPheCysGlnAspCysThrcysLy 768
Qy 748 TAATTGTTGCAACAACCTTGCATCATGATATTGAACGGTTTAAAGCATTAAAGCATGTCT 807
Db 768 sasCysPheAsnAsnLeuAspTyrcysGluValGluArgGluArgAlaIleArgSerCysLe 788
Qy 808 TGGTAGAATCCAGAGCTTCCAGCCAAATAATTGGGAAGGCCCAATTGGCAATGTCAA 867
Db 788 uAspArgAsnProSerAlaPheLysProLysIleThrAlaProAsnSerGlyAspMetAr 808
Qy 868 GCCCAGCACAAAGGGTCAACTGCAGGAGTGCAGGCTGCCTGAAGATTACTCGCA 927
Db 808 g---LeuHisAsnLysGlyCysAsnCysLysArgSerGlyCysLeuLysAsnTyrcysGl 827
Qy 928 GTGCTATGAGGCCCAATATTGTCTCTCTTCTTATTTGCAAAATGCATTTGGTTCACAAATTA 987
Db 827 uCystyrGluAlaLysIleProCysSerSerIleCysLysCysValGlyCysArgAsnMe 847
Qy 988 TGAGAAGCCCAACCAAGACACTACTGATGAGTATGAGTATGAGTATGAGTATGAGTATGAG 1047
Db 847 tGluAspArgProAsp-----ValAspMetAspSerLeuAspGlyLeuMetGl 863
Qy 1048 TTTGAGAGCGACCATTTACCTGCACCAACCAAAATTTTTCAGGACTTCCAGATTTCAGTCA 1107
Db 863 yValGluGlyGln-----LysLysAspLysAlaLysAsnLysGlnLeuAs 878
Qy 1108 CGATAGGGGGCTTCCTCATCATCTCTCGGAGGTGGTGGAGGCCACATCGCCTGCCT 1167
Db 878 nGluAsnArgAlaAsnIleTyrcysPheThrAspValIleGluAlaThrIleMetCysMe 898
Qy 1168 GCTTGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1227
Db 898 tileSerArgIleValMetHisGluLysGlnAsnValAlaValGluAspMetGluArgGl 918
Qy 1228 GATCCTGGAGGAATTTGAGAGGTGCTTATCATCAGATTCTC 1267
Db 918 uValMetGluGluMetGlyGluSerLeuThrGlnIleIle 931
RESULT 7
AAW83392
ID AAW83392 standard; protein; 438 AA.
XX AC AAW83392;
XX DT 29-MAR-1999 (first entry)
XX DE Caenorhabditis elegans synMuv protein LIN-54.
XX KW LIN-54; synthetic multivulvar; synMuv; signal transduction;
XX KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
XX KW cell proliferation; gene therapy.
XX OS Caenorhabditis elegans.
XX PN W09854299-A1.
XX PD 03-DEC-1998.
```

```
XX 28-MAY-1998; 98WO-US11043.
XX PF
XX 28-MAY-1997; 97US-0047996.
XX PR
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX PI Ceol C, Horvitz HR, Lu X;
XX DR WPI; 1999-045362/04.
XX DR N-PSDB; AAV72865.
XX PT Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
XX PT useful for treating diseases associated with altered levels of cell
XX PT proliferation, e.g. carcinomas
XX PS Claim 7; Fig 13; 70pp; English.
XX CC This is the amino acid sequence of LIN-54, a novel protein of
XX CC Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
XX CC synthetic multivulvar (synMuv) gene involved in cell fate and cell
XX CC proliferation, and is part of a pathway that may be used as a
XX CC genetic and biochemical model system for tumour suppression and
XX CC cancer in mammals. synMuv pathway genes and proteins may be used
XX CC to identify genes which are part of the mammalian pathway and
XX CC to identify genes, proteins and therapeutic compounds which
XX CC modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
XX CC C. elegans synMuv polypeptides selected from LIN-37, -35, -55, -52,
XX CC -53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
XX CC (1) nucleic acids (see AAV72866-67) encoding the mouse and human
XX CC homologues of C. elegans LIN-54; (2) vectors containing the nucleic
XX CC acids; (3) transgenic cells; (4) a pure mammalian synMuv
XX CC polypeptide, and (5) an antibody which binds to a synMuv family
XX CC protein. The synMuv nucleic acids and polypeptides can be used to
XX CC diagnose and treat, especially by gene therapy, conditions
XX CC involving altered levels of cell proliferation, e.g. synMuv-
XX CC associated carcinomas.
XX SQ Sequence 438 AA;
Alignment Scores:
Pred. No.: 1,14e-33 Length: 438
Score: 441.00 Matches: 93
Percent Similarity: 55.31% Conservative: 32
Best Local Similarity: 41.15% Mismatches: 61
Query Match: 11.53% Indels: 40
DB: 20 Gaps: 9
US-09-743-237-3 (1-2134) x AAW83392 (1-438)
Qy 698 TACTGTGACTGCTTTCGAGTGGGAGCTTTTGCACCAACTGCAATTTGTAATAATTGTTGC 757
Db 191 TyrCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHis 210
Qy 758 AACAACTTGATCATGATATTTGAACGGTTTAAAGCATTAAAGCATGTCTTTGGTAGAAAT 817
Db 211 AsnAsnIleGluTyrcysPheThrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 230
Qy 818 CCAGAAGCTTTCAGCCCAAAATTTGGG-----AAGGCCCAATTTGGCAATGTCACGCC 871
Db 231 ProAsnAlaPheLysProLysIleGlyIleAlaArgGlyGlyIleThrAspIleGluArg 250
Qy 872 CAGCACAAACAAAGGTGCAACTGCAGGAGTGCAGCTGCCTGGAAGAATTTACTGCGAGTGC 931
Db 251 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrcysGluCys 270
Qy 932 TATGAGGCCCAATATTGTTCTTCTTATTTGCAAAATGCATTTGTTGCAAAAT----- 985
Db 271 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 290
Qy 986 -----TATGAGAAGC-----CCAGAACGAAGACACTA 1015
Db 291 ThrTyrArgMetThrArgTyrcysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeu 310
```


Db 42 CysTyrGlnSerMetAlaIleCysThrLysPheCysArgCysValGlyCysArgAsnThr 61
QY 989 GAA-----GAAGCCCAACGA 1006
Db 62 GluValArgGluLeuValAspProAsnSerValAlaLysAsnSerSerAlaValLysArg 81
QY 1006 ----- 1006
Db 82 GlnLysAlaAlaMetSerAlaLysAlaAlaAlaAlaAlaAlaAlaGlyLeuAsp 101
QY 1007 -----AAGACACTA 1015
Db 102 ValGlnGlyLysAlaLeuGlnValAlaAlaSerThrLeuAlaLeuProGlyLysAlaLeu 121
QY 1016 ATGAGCATGCCAACTACATGCAGACTGGAGT-----TTGGAAGGCGACCATAC 1066
Db 122 MetThrProProLysTyrThrLeuValAlaGlyLysProMetAlaSerSerHisIle 141
QY 1067 CTGCCA-----CCACGAAATTTTCAGGACTTCCAGATTCAGT----- 1105
Db 142 AsnProIleProIleSerArgProIleAlaThrAlaAlaThrProAlaAlaValLys 161
QY 1106 -----CACCATAGG 1114
Db 162 GlnProAlaGluProProMetProValAsnLeuIleIleProValArgHisAspArg 181
QY 1115 CGGCCTCTCATGCTCTCTGCGAGGTGGTGGAGGCGACATGCGCTGCTGCT 1174
Db 182 ArgAspArgAsnLeuPheValGlnProValAsnAlaAlaLeuLeuGluCysMetLeuIle 201
QY 1175 CAGGAGAGAGCGCGAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCTCG 1234
Db 202 GlnAlaThrGluAlaGluGlnLeuGluLeuAsnGluLeuGlnValCysGlnLeuValLeu 221
QY 1235 GAGGAATTGGAGGTGCTTATCAGATTCTC 1267
Db 222 GluGluPheMetArgGlyTyrLysAsnIleLeu 232

RESULT 11
ID ABG30203 standard; Protein; 4561 AA.
XX ABG30203;
AC ABG30203;
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #30194.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS94390.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID No 60562; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4561 AA;

Alignment Scores: 7.41e-05 Length: 4561
Pred. No.: 149.00 Matches: 157
Score: 28.98% Conservativity: 54
Best Local Similarity: 21.57% Mismatches: 205
Query Match: 3.90% Indels: 312
DB: 22 Gaps: 45
US-09-743-237-3 (1-2134) x ABG30203 (1-4561)
QY 8 GGTCAAGCGAAGCTCGCGGGGCGACAGCGCGGGAGCTCT-----CGG 58
Db 1647 GlySerGlyArgCysSerAlaGlyArgProGlyLeuProGlyProAlaProLysAla 1666
QY 59 GGAGTACCCCGGATCCAG----- 79
Db 1667 AlaLeuProArgGluGlyAlaLeuAlaGlyAlaValProAspSerAlaProAlaLeu 1686
QY 80 -----GCTCAGCGCGCTGGAGGAGCTCGGCTCTCGAGGCCCCCGACCGCC 127
Db 1687 GlyIleHisProAlaGlnGluMetProProArgSerProAlaAlaProLeuTyrAla 1706
QY 128 CGCCTGCACAGTGCAC-----TCCGTGCTCT----- 153
Db 1707 GlnIleGlu--CysThrGlyPheCysAlaProGlyCysThrCysProGlyLeuPheL 1726
QY 154 -----CGCTGTACCGCGCACCGCGAGCCCGGGGT----- 186
Db 1726 euHisAsnAlaSerCysLeuProArgSerGlnCys---ProCysGlnLeuHisGlyGlnL 1745
QY 187 ---TTTGCCCTGGGCGCGCTGGG-----TCTGCGAAGGAGCTCCACCGCGGG 235
Db 1745 eutyTyrAlaSerGlyAlaMetAlaArgLeuAspSerCysAsnAsnCysCysAsnValGlyI 1765
QY 236 TCCGCATGATCCAGTTGAAATCAAGGTAAGCAGGTGTACTTACT----- 280
Db 1765 leArgArg-----ArgPhe-ArgAlaGlyThrAlaProProAlaAlaPhe 1779
QY 281 -----ACAAGTATATCCGGAAGAGCAACTTTGCAGATCTT 319
Db 1780 GlyGlyAlaGluCysGlnGlyProThrMetGluAlaGluPheCysSerLeuArgProCys 1799
QY 320 CTTGTCTCAGGAATCCTGTTC---AAGTTCCTCATGGTCCCGAGAACTAGAGATCCCTCC 376
Db 1800 ProAlaSerTrpGlyCysHisArgValProCysThrGlyGluLeuGlu----- 1816

```
QY 377 TGCTGTTCTCTTAAGAAGATTCCACCCCAATGGTGATATGCCCAATTAAGAGGGGCACA 436
Db 1817 -----GlnArgProLeuMetValSerGlnIle-----
QY 437 CAAATGCTATGTATAGACAATTTCTAGAACAGAGAACTAAAGCACTCCATTTGGTTCCCT 496
Db 1826 -----LeuGluAla-----
QY 497 CAGTATCAAGATCAAAATAATTAATCTACAGTCAGATGCTCCCTAAACCAATGACTGCT--- 553
Db 1829 -----GlnAspGlnGlyValAlaProValSerProGlySerSerProThrAlaAlaPro 1846
QY 554 ---TTAGTAGGAGCATTTTGGCCAGCATCAACAAAATAATCTC---ATTACACAACA 607
Db 1847 GlyLeuValArgValSerProArgGlyHisProGlyProLeuGlyAlaSerValGln 1866
QY 608 CTTGAGGAGCCTTACCATTACCGTAGTCAACGGGTCTCTTCCCTCGGATCA----- 661
Db 1867 ProProValAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAlaGlyTyr 1886
QY 661 -----
Db 1887 GlyProTrpGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArg 1906
QY 662 -----ACTCTTCAGGACACACAAAATAACTTTGCTGGTACTGTGAC---TGCTTT 712
Db 1907 ThrArgAlaCysAspGlnProProGlnGlyLeuGlyAspTyrCysGluGlyProArg 1926
QY 713 GCCAGTGGGACTTTTGC-----AACACTGCAT----- 742
Db 1927 AlaGlnGlyGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGlu 1946
QY 743 -----TGTAATAATTTGTTGCCAACAACTTGTCATCAT 772
Db 1947 GlyAlaGluTyrSerProCysGlyProProCysProArgSerCysAspAspLeuValHis 1966
QY 773 GATATTGAACGGTTAAAGCCATTAAAGCATGTCTGGTAGA---AATCCAGAGCTTTC 829
Db 1967 -----CysValTrpArgCysGlnProGlyCysTyr 1976
QY 830 CAGCCAAAATTGGAGAGGCCAATTGGGCAAT-----GTCAAGCCC 871
Db 1977 CysPro-----ProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnPro 1993
QY 872 CAGCAACAAGGGTGCACACTGC----- 895
Db 1994 GlyHis-----CysSerCysLeuAspLeuLeuThrGlyGlnArgHisProGly 2010
QY 896 -----AGGAGGTGAGCTGCCTGAAGAAATTAAGTCTGAGTCTATGAGGCCCAA 943
Db 2011 AlaArgLeuAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGluGlyArg 2028
QY 944 ATTATGTGTTCTTCTATTGTCAAATGATGTTGTTGCAAAAATTTATGAAGAAAGCCAGAA 1003
Db 2029 LeuAsnCysThrAspLeu-----ProCysProAspCys----- 2039
QY 1004 CGAAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGTTTGAAGGCAGCCAT 1063
Db 2040 -----GlyGlyGlyGlnSerLeuHis 2046
QY 1064 TACTGCCCAACGAAATTTTCAGACATTCGAAGATTACGTACGATAGGCGGCTTCC 1123
Db 2047 -----ProCysGly-----GlnProCysProArgSerCysGlnAspLeuSerProGly 2062
QY 1124 TCA---TGATCTCTCTGGGAGGTG---GTGGAGGCCACATGCGCTGCTGTGCTCAG 1177
Db 2063 SerValCysGlnProGlySerValGlyCysGlnProThrCysGlyCysProLeuGlyGln 2082
QY 1178 -----GGAGAAGAGGCCGAGAAAGAACACTGTCTCAAGTGCCTGGCAGACGAG 1225
Db 2083 LeuSerGlnAspGlyLeuCysValProProAlaHisCys----- 2095
QY 1226 ATGATCTGGAGGAATTTGGAAGGTCTTATCACAG-----ATTCCTCCAC 1270
```

```
Db 2096 -----ArgCysGlnTyrGlnProGlyAlaMetGlyIleProGlu 2108
QY 1271 ACTGAGTTTAAATCTAAGGATTGAAATGGAG-----TAGAGTATAAAG----- 1315
Db 2109 AsnGlnSerArgSerAlaGlySerArgPheSerSerTrpGluSerLeuGluProGlyGlu 2128
QY 1316 -----TGTGAATGCATGTTGATTTTGTCTTGTAGCTAGAAAT 1351
Db 2129 ValValThrGlyProCysAspAsnCysThrCysValAlaGlyIle----- 2143
QY 1352 CTCTAGTTTAGAAGGATGTTTAGGGGAACATGAGGCTGCTGTGACAGCAACACAGGC 1411
Db 2144 -----LeuGlnCysGlnGluValProAsp 2151
QY 1412 TCCCT-----GCATCCCTG 1426
Db 2152 CysProAspProGlyValTrpSerSerTrpGlyProTrpGluAspCysSerValSerCys 2171
QY 1427 GGGCCAGGGAGTTTACTTCAGAGCTCTCTGAAGATGTGGCAAC---CCATGCCCTTTTC 1483
Db 2172 GlyGlyGlyGluGlnLeuArgSerArg---ArgCysAlaArgProCysPro----- 2188
QY 1484 TGAGGAGGTGCATGGCTGAGCATTTGTTCTGGCCAGAGAGAGAGCTTGGTTCC 1543
Db 2189 ---GlyProAla-ArgGlnSerArgThrCysSerThrGln----- 2200
QY 1544 ATAGTCTGGGACAGTCTCTGCAGGCGCGGAGGCGAGCAGCAGGC----- 1589
Db 2201 -----ValCysArgGluAlaGlyCysProAlaGlyArgLeuTyrArgG1 2215
QY 1590 -----CCTCGGAGAGCTCACTGCTGCTCTCTCTCAGAGAA-----TG 1633
Db 2215 uCysGlnProGlyGluGlyCysProPheSerCysAlaHisValThrGlnGlnValGlyCy 1633
QY 1634 TTCTCTGGAGGCTGC 1649
Db 2235 sPheSerGluGlyCys 2240
RESULT 12
ABG21064
ID ABG21064 standard; Protein; 9222 AA.
AC ABG21064;
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #21055.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS85251.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
```

PT biodiversity -
 XX Claim 20; SEQ ID NO 51423; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 9222 AA;

Alignment Scores:

Pred. No.:	0.000101	Length:	9222
Score:	149.00	Matches:	157
Percent Similarity:	28.98%	Conservative:	54
Best Local Similarity:	21.57%	Mismatches:	205
Query Match:	3.90%	Indels:	312
DB:	22	Gaps:	45

US-09-743-237-3 (1-2134) x ABG21064 (1-9222)

QY	8	GGTCAAGCGAAGCTCGGGGGGCGACAGCGCGGGAGTCCT-----CGG	58
DB	2371	GlySerGlyArgCysSerAlaGlyArgProGlyLeuProGlyProAlaProLysAla	2390
QY	59	GGAGTACCCCGGATCCACAG-----	79
DB	2391	AlaLeuProArgGluGlyAlaLeuAlaGlyAlaValProAspSerAlaProAlaLeu	2410
QY	80	-----GCTCAGCGCGCTGGAGGACGTCGCGCTCTCGAGCGCCCGCAGCGCC	127
DB	2411	GlyLeuHisProAlaGlnGluMetProProArgArgSerProAlaAlaProLeuTyrAla	2430
QY	128	CGCTGCAACCTGCACCT-----TCCTGTCTCT-----	153
DB	2431	GlnIleGlu--CysThrGlyPheCysAlaProGlyCysThrCysProProGlyLeuPheLe	2450
QY	154	-----CGCTGTACCCGCCACCGCGCGCGCGGTG-----	186
DB	2450	euHisAsnAlaSerCysLeuProArgSerGlnCys---ProCysGlnLeuHisGlyGlnL	2469
QY	187	-----TTTGGCCCTGGCGCGCTGGG-----TCTTGGGAAGGAGCTCCACCGCGCG	235
DB	2469	euTyrAlaSerGlyAlaMetAlaArgLeuAspSerCysAsnAsnCysCysAsnValGlyI	2489
QY	236	TCCGCATGATCCAGTTGAATCAAGGTAAGCAGGTGCTACTACT-----	280
DB	2489	leArgArg-----ArgPhe-ArgAlaGlyThrAlaProProAlaAlaPhe	2503
QY	281	-----ACAAGTAATAATCCGGAAGAAGCAACTTTTCAGAAATCTT	319
DB	2504	GlyGlyAlaGluCysGlnGlyProThrMetGluAlaGluPheCysSerLeuArgProCys	2523
QY	320	CTTCGCTCAGGAATCCTGTTC-----AGTTCCCATGGTCCCGAGGAAGTACAGATGCCTCC	376
DB	2524	ProAlaSerTrpGlyCysCysHisArgValProCysThrGlyGluLeuGlu-----	2540

QY	377	TGCTGTTCTCTTAAGAAAGATTCCCAACCAATGCTGATATCCCAATTTGAAGGGGGCACA	436
DB	2541	-----GlnArgProLeuMetValSerGlnIle-----	2549
QY	437	CAAAATGCTATGTATAGACAATTTCTAGAACAAAGAGAACTTAAAGAGCACTCCATTTGGTTCCT	496
DB	2550	-----LeuGluAla-----	2552
QY	497	CAGTATCAAGATCAAAATAATATTCTACAGTCAGATGTCCTAAACCAATGACTGCT---	553
DB	2553	-----GlnAspGlnGlyValAlaProValSerProGlySerSerProThrAlaAlaPro	2570
QY	554	---TTAGTAGGAGCATTTTCCAGCATCAACAAAATTAATATCTC---ATTACACAACAA	607
DB	2571	GlyLeuValArgArgValSerProArgGlyHisProGlyProLeuGlyAlaSerValGln	2590
QY	608	CTTGAGGAGCCTTACCATCGTAGTCAACGGGTCTGCTTCCCTCCGCGATCA-----	661
DB	2591	ProProValAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAlaGlyGlyTrp	2610
QY	661	-----	661
DB	2611	GlyProTrpGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArg	2630
QY	662	-----ACTCTTCAGGAGCACCAAAATAACTTTGGCTGGGTACTGTGAC---TGCTTT	712
DB	2631	ThrArgAlaCysAspGlnProProGlnGlyLeuGlyAspTyrCysGluGlyProArg	2650
QY	713	GCCAGTGGGGCTTTTTCG-----AACAACTGCAAT-----	742
DB	2651	AlaGlnGlyGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGlu	2670
QY	743	-----TGTAAATAATTCTTGCACAACTTGCATCAT	772
DB	2671	GlyAlaGluTyrSerProCysGlyProProCysProArgSerCysAspLeuValHis	2690
QY	773	GATATTGAACGGTTTAAAGCCATTGAAGCATGCTCTGTGTAGA---AATCCAGAAAGCTTTC	829
DB	2691	-----CysValTrpArgCysGlnProGlyCysTrp	2700
QY	830	CAGCAGAAAAATTGGAGGGCAATTTGGGCAAT-----GTCAAGCCC	871
DB	2701	CysPro-----ProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnPro	2717
QY	872	CAGCACAAACAGGTCGCACTGC-----	895
DB	2718	GlyHis-----CysSerCysLeuAspLeuLeuThrGlyGlnArgHisHisProGly	2734
QY	896	-----AGGAGGTGAGCTGCTGAAGAAATTAATCTCGAGTGTGATGAGGCCCAA	943
DB	2735	AlaArgLeuAlaAArgProAspGlyCys-----AsnHisCysThrCysLeuGluGlyArg	2752
QY	944	ATTATGTCTCTCTTATTGCAATGTCATGTTGCAAAAATTATGAAGAAGCCAGAA	1003
DB	2753	LeuAsnCysThrAspLeu---ProCysProAspCys-----	2763
QY	1004	CGAAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGGAGGCGCAT	1063
DB	2764	-----GlyGlyGlyGlnSerLeuHis	2770
QY	1064	TACCTGCCACCAACGAAATTTTCAGGACTTCCAGATTTCAGTCAGATGAGCGGCTTCC	1123
DB	2771	-----ProCysGly---GlnProCysProArgSerCysGlnAspLeuSerProGly	2786
QY	1124	TCA---TGCACTCTCTGGGAGGTG---GTGGAGGCCACATCGCTGCTGCTGCTCAG	1177
DB	2787	SerValCysGlnProGlySerValGlyCysGlnProThrCysGlyCysProLeuGlyGln	2806
QY	1178	-----GGAGAGAGCGCGAGAAAGAACACTCTCTCAAGTGCCTGGCAGAGCAG	1225
DB	2807	LeuSerGlnAspGlyLeuCysValProProAlaHisCys-----	2819

QY 1226 ATGATCTGGAGGAATTTGGAAGTGTCTTATCAG-----ATTCTCCAC 1270
 Db 2820 -----ArgCysGlnTrpGlnProGlyAlaMetGlyIleProGlu 2832
 QY 1271 ACTGAGTTAAATCAAGGATTGAAATGAG-----TAGAGTATAAG----- 1315
 Db 2833 AsnGlnSerArgSerAlaGlySerArgPheSerSerTrpGluSerLeuGluProGlyGlu 2852
 QY 1316 -----TGTGAATGCATGTGATTGTTCTTACTGTAGCTAGAAAT 1351
 Db 2853 ValValThrGlyProCysAspAsnCysThrCysValAlaGlyIle----- 2867
 QY 1352 CTCTAGTTAGAAAGGATTTAGGGGAACATGAGGCTGCTGCAGCAACACAGGC 1411
 Db 2868 -----LeuGlnCysGlnGluValProAsp 2875
 QY 1412 TCCCT-----GCATCCCTG 1426
 Db 2876 CysProAspProGlyValTrpSerSerTrpGlyProTrpGluAspCysSerValSerCys 2895
 QY 1427 GGCCCAAGGAGTTACTCAGAGCTCTGAAGATGTGGCAAC---CCATGCCCCCTTTTC 1483
 Db 2896 GlyGlyGlyGluGlnLeuArgSerArg---ArgCysAlaArgProProCysPro----- 2912
 QY 1484 TGAGGAGTGCATGGCTGAGCTGTTCTGTGCGCCAGAGAGAGAGCTGGGTGCC 1543
 Db 2913 ---GlyProAla---ArgGlnSerArgThrCysSerThrGln----- 2924
 QY 1544 ATAGTCTGGGAGAGTGTCTGCGAGCGCGGAGGCGAGCAGCGC----- 1589
 Db 2925 -----ValCysArgGluAlaGlyCysProAlaGlyArgLeuTrpArgG1 2939
 QY 1590 -----CCTGCGAGAGCTCACTGTGCGACTCTTCTCTCAGAGAA-----TG 1633
 Db 2939 uCysGlnProGlyGluGlyCysProPheSerCysAlaHisValThrGlnGlnValGlyCy 2959
 QY 1634 TTCCTCTGGAGGCTGC 1649
 Db 2959 sPheSerGluGlyCys 2964
 RESULT 13
 ABB81598
 ID ABB81598 standard; Protein; 2743 AA.
 AC ABB81598;
 XX
 XX
 19-SEP-2002 (first entry)
 XX
 Human laminin alpha 5 2743 N-terminal amino acid sequence SEQ ID NO:36.
 DE
 XX
 Laminin alpha 5; laminin 10; vulnerary; cell growth; differentiation;
 KW tissue repair development; laminin; healing; vascular tissue;
 KW re-endothelialisation; vascular injury; cell attachment; cell stasis;
 KW proliferation; migration.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..35
 FT /label= signal
 FT Protein 36..2743
 FT /label= laminin_alpha_5
 XX
 PN WO200250111-A2.
 XX
 XX 27-JUN-2002.
 PD
 XX 21-DEC-2001; 2001WO-US51035.
 PF
 XX 21-DEC-2000; 2000US-257449P.
 PR 28-MAR-2001; 2001US-279282P.
 PR 13-NOV-2001; 2001US-0279282.
 XX

(BIOS-) BIOSTRATUM INC.

Tryggvason K, Doi M, Thyboll J;

WPI; 2002-557650/59.

N-PSDB; ABQ72930.

New human laminin-10 proteins, useful for accelerating the healing of vascular tissue, improving the biocompatibility of grafts, or for promoting re-endothelialization at the site of vascular injuries

Disclosure; Page 223-231; 231pp; English.

The present invention describes human laminin alpha 5. Also described is an isolated laminin 10. Laminin 10 has vulnerary activity. Laminins are useful in maintaining cell/tissue phenotype as well as promoting cell growth and differentiation in tissue repair development. Specifically, laminin 10 can be used for accelerating the healing of injuries of vascular tissue, improving the biocompatibility of grafts useful for treating such injuries, for promoting re-endothelialisation at the site of vascular injuries, and promote cell attachment and subsequent cell stasis, proliferation, differentiation, and/or migration. The present sequence represents the 2743 N-terminal amino acid sequence of human laminin alpha 5, which is used in the exemplification of the present invention.

SQ Sequence 2743 AA;

Alignment Scores:

Pred. No.: 0.00373 Length: 2743
 Score: 130.50 Matches: 111
 Percent Similarity: 31.14% Conservative: 55
 Best Local Similarity: 20.83% Mismatches: 198
 Query Match: 3.41% Indels: 170
 DB: Gaps: 28

US-09-743-237-3 (1-2134) x ABB81598 (1-2743)

QY 18 AAGCTCGCGGGGCGACAGCGACGGGGAG-----CTCCTCGGGAGTAC 65

Db 83 LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleArgGlyGlnTyr 102

QY 66 CCCGGGATC-----CCAGAGCTCAGCGCGGTGGAG 95

Db 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleAsp 122

QY 96 GACGTC-----GGCTCTCGAGCGCCCGCGCGCGCTGC 134

Db 123 GlyThrGluArgTrpTrpGlnSerProProLeuSerArgGlyLeuGluTyrAsnGluVal 142

QY 135 AACGTG-----CACTTCTCTGCTCGCTGCTACCCGCGCAC 170

Db 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 162

QY 171 CGCAGCGCGGGTGTGTTGCGCGCGCTGGTCTCGGAGGAGGCTCCACCC 230

Db 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe 178

QY 231 GGGCGTCCGATGATCCAGTTGAAATCAAGTAAAGCAGGTGGTACTACTACAAGTAATA 290

Db 179 GlyArgThrTrpGlnPro-----TrpGlnPhePhe 188

QY 291 ATCCGGAAGAACAACTTTGCAAGATCTTTGCTCAGGAATCTCTGTTGCAAGTTCCCAT 350

Db 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199

QY 351 GGTCCCGAGAACTA-----GAGGATGCTCTGCTGCTTCTTCTTAAAGAA 394

Db 200 GlyProGlnThrLeuGluArgIleThrArgAspAspAlaIleCysThrThrGluTyr 219

QY 395 GATTCACCAACCAATGGTGATATGCCAATTTGAAAGGGGCGACACAATGCTGTATAGAC 454

Db 220 SerArg-----IleValProLeuGluAsnGlyGluIleValSerLeuVal 235

RESULT 14
ABB81588
ID ABB81588 standard; Protein; 3695 AA.
XX AC ABB81588;
XX DT 19-SEP-2002 (first entry)
XX XX Human laminin alpha 5 protein SEQ ID NO:2.
XX KW Laminin alpha 5; laminin 10; vulnerin; cell growth; differentiation; tissue repair development; laminin; healing; vascular tissue; re-endothelialisation; vascular injury; cell attachment; cell stasis; proliferation; migration.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..35
FT /label= signal
FT Protein 36..3695
FT /label= laminin_alpha_5
XX PN WO200250111-A2.
XX PD 27-JUN-2002.
XX PF 21-DEC-2001; 2001WO-US51035.
XX PR 21-DEC-2000; 2000US-257449P.
XX PR 28-MAR-2001; 2001US-279282P.
XX PR 13-NOV-2001; 2001US-0279282.
XX PA (BIOS-) BIOSTRATUM INC.
XX PI Tryggvason K, Doi M, Thyboll J;
XX DR WPI; 2002-557650/59.
XX DR N-PSDB; ABQ72906.
XX PT New human laminin-10 proteins, useful for accelerating the healing of vascular tissue, improving the biocompatibility of grafts, or for promoting re-endothelialization at the site of vascular injuries - Claim 5; Page 68-79; 231pp; English.
XX CC The present sequence represents human laminin alpha 5. Also described is an isolated laminin 10. Laminin 10 has vulnery activity. Laminins are useful in maintaining cell/tissue phenotype as well as promoting cell growth and differentiation in tissue repair development. Specifically, laminin 10 can be used for accelerating the healing injuries of vascular tissue, improving the biocompatibility of grafts useful for treating such injuries, for promoting re-endothelialisation at the site of vascular injuries, and promote cell attachment and subsequent cell stasis, proliferation, differentiation, and/or migration.
XX SQ Sequence, 3695 AA;
Alignment Scores:
Pred. No.: 0.00424 Length: 3695
Score: 130.50 Matches: 111
Percent Similarity: 31.14% Conservative: 55
Best Local Similarity: 20.83% Mismatches: 198
Query Match: 3.41% Indels: 170
DB: 23 Gaps: 28
US-09-743-237-3 (1-2134) x ABB81588 (1-3695)
Qy 18 AAGCTCCGGGGCGACACGACGGGGGAG-----CTCCTCGGGAGTAC 65
Db 83 LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrfileargGlyGIntyr 102

```

QY 66 CCGGGATC-----CCAGAGCTCAGCGCGCTGGAG 95
Db 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleAsp 122
QY 96 GACGTC-----CGGCTCTGAGCGCCCGCAGCGCGCCGCTGC 134
Db 123 GlyThrGluArgTrpTrpGlnSerProProLeuSerArgGlyLeuGluTrpAsnGluVal 142
QY 135 AACGGT-----CACITCTCTGCTCGCTCTACCCCGCGCAC 170
Db 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaIleValLeuIleLysPheAla 162
QY 171 CGCAGCCGCGGGTGTGTTTGGCCCTGGCGGCTGGTCCGCGAGGAGGCTCCACCC 230
Db 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe 178
QY 231 GGGCGTCGCATGATCCAGTTGAATCAAGTAAGCAGGTGGTACTACTACAGTAATA 290
Db 179 GlyArgThrTyTrpGlnPro-----TrpGlnPhePhe 188
QY 291 ATCCGGAAGAAGCAACTTTGAGAGATCTTCTTGCTCAGGAATCCTGTTGCAAGTTCCCAT 350
Db 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199
QY 351 GTTCCCGGAAGT-----GAGGATGCTCTCTGCTCTCTTTAAAGAA 394
Db 200 GlyProGlnThrLeuGluArgIleThrArgAspAlaAlaIleCysThrThrGluTrp 219
QY 395 GATTCCACCCCAATGGTATGCAATTTGAAGGGGCGCACAAATCTGTATGATAGAC 454
Db 220 SerArg-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal 235
QY 455 AATTCTAGAACAGAGAACTAAAGCACTCCATTGGTTGGTCTCAGTATCAAGATCAAAAT 514
Db 236 AsnGlyArgProGlyAlaMet----- 242
QY 515 AATTATCTACAGTACAGTGTCTTAACCAATAGTACTGCTTTAGTAGG-----AGA 565
Db 243 AsnPheSerTyTrpSerProLeuArgGluPheThrLysAlaThrAsnValArgLeuArg 262
QY 566 TTTTTCGCCAGCATCAACAAATTA---AATCTCATTAACAACACTTGGAGGCCCTTA 622
Db 263 PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp----- 280
QY 623 CCATCGGTAGTCAACGGGTCTGTTTCCCTCGGATCAACTCTCCAGGACCCACAAA 682
Db 281 ProThrValThrArgArgTyTrp-----SerIleLysAsp 293
QY 683 ATAACTTTGGTGGTACTGTGACTGCTTGGCAGTGGGACTTTTGC----- 730
Db 294 IleSerIleGlyArgCysValCysHisGlyHisAlaAspAlaCysAspAlaLysAsp 313
QY 731 -----AACNACTGC-----AATGT 745
Db 314 ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyGlyThrCys 333
QY 746 AATAATTGTCACAACTGCATCATGATATTGAAGCGTTTAAAGCCATTAAAGCA--- 802
Db 334 AspArgCysCysProGlyPheAsnGln-----GlnProTrpLysProAlaThrAlaAsn 351
QY 803 -----TGCTTGTGTAGA-----AAT 817
Db 352 SerAlaAsnGluCysGlnSerCysAsnCysTyTrpGlyHisAlaThrAspCysTyTrpAsp 371
QY 818 CCAGAA-----GCITTCAGCCAAAATTTGGAGGGCCCAATTTGGCCAAAT 862
Db 372 ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyTrpGlnGlyGly 391
QY 863 GTC-----AAGCCCCAGCAGCAACAAAGGTCGCAATTCAGGAGGTTCAGGCTGCTG 913
Db 392 ValCysIleAspCysGlnHisThrAlaGlyValAsnCysGluArg-----CysLeu 409
QY 914 AAGAAATTACTCGAGTGTATGAGGCCCAAAATATGTCTTCTTATTTGCAATGCAT 973

```

```

Db 410 ProGlyPheTyArgSerProAsnHisProLeuAspSerProHisValCysArgArgCys 429
QY 974 GTTTCACAAAT-----TATGAGAAAGCCCAAGACGAAAGACACTA 1015
Db 430 AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys 447
QY 1016 ATGAGCATGCCAAACTAC-----ATGCAGACTGGAGGTTTGGAAAGC 1057
Db 448 TyrCysArgProAsnPheSerGlyLysArgCysAspValCysAlaGluGlyPheThrGly 467
QY 1058 AGCCATTACCTGCCCAACCAAGAAATTTTCAGGACTTCCAAGATTCAGTCACGATAGCGG 1117
Db 468 PheProSerCysTyTrpProThr-----ProSerSerAsnAspThrArg 482
QY 1118 CTTTCCCTCATCATCTCTGGAGGTGGTGGAGGCCACATCGCTGCTGCTGCTCTCAG 1177
Db 483 GluGlnValLeuProAlaGlyGlnIleValAsnCysAspCysSerAlaAlaGlyThrGln 502
QY 1178 GGAGAAGAGCGCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGATCCTGGAG 1237
Db 503 GlyAsnAlaCysArgLysAsp-----Pro 510
QY 1238 GAATTTGGAAGGTGCTTATCA-----CAGATTCCTCCACACTGAGTTTAAATCT 1285
Db 511 ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla 530
QY 1286 AAGGATTTGAAATGGAGTAGATATAAGTGTGAATGC 1324
Db 531 ProGlyPheTyArgGlyProGlyCysGlnProCysGlnCys 543
RESULT 15
AAE17310
ID AAE17310 standard; Protein: 3696 AA.
XX
AC AAE17310;
XX
DT 18-APR-2002 (first entry)
XX
DE Human laminin alpha protein, sbg417005LAMININ_ALPHA #2.
XX
KW Human; therapy; wound healing disorder; vaccine; cancer; infection;
KW autoimmune disorder; haematopoietic disorder; inflammation; arthritis;
KW Parkinson's disease; Huntington's chorea; schizophrenia; antiarrhythmic;
KW multiple sclerosis; Alzheimer's disease; analgesic; cardiant; asthma;
KW ischaemia; stroke; AIDS; bone disease; atherosclerosis; brain disorder;
KW depression; cardiovascular disease; myocardial infarction; renal failure;
KW respiratory disease; liver disorder; Fanconi's syndrome; spleen disorder;
KW type II diabetes mellitus; skeletal muscle disorder; immunosuppressive;
KW hypersplenism; renal disease; hypoglycaemia; gastrointestinal disease;
KW noctropic; cirrhosis; Hodgkin's disease; neuroleptic; antiinflammatory;
KW haemostatic; vulnery; anticonvulsant; antirheumatic; neuroprotective;
KW nephrotropic; hypotensive; vasotonic; cytostatic; cerebroprotective;
KW allergy; laminin alpha protein.
XX
OS Homo sapiens.
XX
WO200198342-A1.
XX
PN 27-DEC-2001.
XX
PD 22-JUN-2001; 2001WO-US19929.
XX
PF 22-JUN-2000; 2000US-213156P.
XX
PR 22-JUN-2000; 2000US-213161P.
XX
PA (SMK ) SMITHKLINE BEECHAM CORP.
PA (SMK ) SMITHKLINE BEECHAM PLC.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Agarwal P, Cogswell JP, Kabnic KS, Lai Y, Martensen SA;
PI Murdock PR, Smith RF, Strum JC, Xiang Z, Xie Q, Rizni SK;
XX

```

DR WPI: 2002-139783/18.
 XX N-PSDB; AAD27805.
 PT Novel secreted and membrane-associated polypeptides and polynucleotides
 PT useful for preventing, ameliorating or correcting dysfunction or
 PT disease including diabetes, cancer, hypertension and growth
 PT abnormalities
 XX
 Claim 1: Page 115-122; 138pp; English.
 PS
 XX The invention relates to secreted and membrane-associated polypeptides
 CC and polynucleotides. The sequences of the invention are useful in
 CC diagnostic assays for detecting diseases associated with inappropriate
 CC activity or levels of these polynucleotides, and in identifying their
 CC agonists and antagonists that are potentially useful in therapy. The
 CC sequences of the invention are useful as vaccines for inducing
 CC immunological response. The sequences of the invention are useful for
 CC treating cancers, infections, autoimmune disorders, hematopoietic
 CC disorders, wound healing disorders, cholesterol ester storage disease,
 CC inflammation, congenital muscular dystrophy, junctional epidermolysis
 CC bullosa, Parkinson's disease, Huntington's chorea, multiple sclerosis,
 CC viral and bacterial infections, Alzheimer's disease, asthma, arthritis,
 CC allergies, schizophrenia, sbg442445PROA-associated disorders,
 CC septicemia, psoriasis, inflammatory bowel disease, transplant rejection,
 CC graft versus host disease, ischemia, stroke, acute respiratory disease
 CC syndrome, restenosis, brain injury, AIDS, bone diseases, atherosclerosis,
 CC brain disorders including paraspranuclear palsy, myotonic dystrophy,
 CC depression, anxiety disorders and sleep disorders, cardiovascular
 CC diseases including congestive heart failure and myocardial infarction,
 CC respiratory diseases including chronic obstructive pulmonary disease,
 CC acute bronchitis and adult respiratory distress syndrome, liver disorders
 CC including hypercholesterolemia, hypertriglyceridemia, cirrhosis, viral
 CC and non-viral hepatitis, type II diabetes mellitus, renal disease
 CC including acute and chronic renal failure, glomerulonephritis, Fanconi's
 CC syndrome, cystinuria, skeletal muscle disorders including hypoglycemia
 CC and tendinitis, gastrointestinal diseases including intestinal
 CC obstruction and tropical sprue, spleen disorders including hypersplenism,
 CC Hodgkin's disease and malignant lymphoma, testicular cancer, male
 CC reproductive diseases including low testosterone and male infertility.
 CC The present sequence is human laminin alpha protein.
 XX
 SQ Sequence 3696 AA;

Alignment Scores:
 Pred. No.: 0.00424 Length: 3696
 Score: 130.50 Matches: 111
 Percent Similarity: 31.14% Conservative: 55
 Best Local Similarity: 20.83% Mismatches: 198
 Query Match: 3.41% Indels: 170
 DB: 23 Gaps: 28

US-09-743-237-3 (1-2134) x AAEL17310 (1-3696)

QY	18	AGCTCGCGGGGACACGACGCGGGAG-----CTCCTCGGGAGTAC	65
Db	83	LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleArgGlyGlnTyr	102
QY	66	CCCGGATC-----CCAGCTCAGCGCGCTGGAG	95
Db	103	CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleAsp	122
QY	96	GACGTC-----GGCTCTCGAGCGCCGCGCGCGCTGC	134
Db	123	GlyThrGluArgTrpGlnSerProProLeuSerArgGlyLeuGlnTyrAsnGluVal	142
QY	135	ACGGT-----CACTTCCTCTCGCTCGCTACCGCGCAC	170
Db	143	AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla	162
QY	171	CGCAGCCCGCGGTGTTTGGCCCTGGGCGCTGGTCTCGAAGGAGGCTCCACCC	230
Db	163	AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe	178
QY	231	GGCGCTCGCATGATCCAGTTGAAATCAAGGTAAGCAGGTGCTACTACTACAAGTAATA	290
Db	179	GlyArgThrGlnPro-----TrpGlnPhePhe	188
QY	291	ATCCGGAAGCAAGCAACTTTTGCAGATCTTCTTCTCAGGAATCCTGTTCGAAGTCCCAT	350
Db	189	AlaSerSerLysArgAspCys-----LeuGluArgPhe	199
QY	351	GGTCCCGGAACTA-----GAGGATGCTCTCTCTCTCTCTTAAGAA	394
Db	200	GlyProGlnThrLeuGluArgIleThrArgAspAlaAlaIleCysThrThrGluTyr	219
QY	395	GATTCACCAACCAATGTGATATGCCAATTCGAAGGGGACACAAATGCTATATAGAC	454
Db	220	SerArg-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal	235
QY	455	AATTCTAGAACAGAACTAAAGCACTCCATTTGGTCTCCTCAGTATCAAGATCAAAAT	514
Db	236	AsnGlyArgProGlyAlaMet-----	242
QY	515	AATTATCTACAGTCAGATGCTCCCTAACCAATCACTGCITTAGTGGG-----AGA	565
Db	243	AsnPheSerTyrSerProLeuLeuArgGluPheThrLysAlaThrAsnValArgLeuArg	262
QY	566	TTTTTCCAGCATCAACAAATTA---AATCTATTACACAACTTGAGGGAGCCTTA	622
Db	263	PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp-----	280
QY	623	CCATCGGTAGTCAACGGGTCTGCTTCCCTCGGGATCAACTTCCAGGACACCACAAA	682
Db	281	ProThrValThrArgArgTyrTyr-----SerIleLysAsp	293
QY	683	ATAACTTTGGCTGGGTACTGTGACTCTTCCAGTGGGACTTTTGC-----	730
Db	294	IleSerIleGlyArgCysValCysHisGlyHisAlaAspAlaLysAsp	313
QY	731	-----AACAACTGC-----AATGT	745
Db	314	ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyGlyThrCys	333
QY	746	AATAATTGTGCAACAACTTGCATCATGATATGTAAGGTTTAAAGCCATTAAAGCA---	802
Db	334	AspArgCysCysProGlyPheAsnGln-----GlnProTrpLysProAlaThrAlaAsn	351
QY	803	-----TGCTCTGGTAGA-----AAT	817
Db	352	SerAlaAsnGluCysGlnSerCysAsnCysTyrGlyHisAlaThrAspCysTyrTyrAsp	371
QY	818	CCAGAA-----GCTTTCCAGCCAAAATTTGGGAAGGGCCCAATTTGGGCAAT	862
Db	372	ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyrGlnGlyGly	391
QY	863	GTC-----AAGCCCCAGCACACAAGGTGCACTGCAGGAGGTGAGGTGCTGCTG	913
Db	392	ValCysIleAspCysGlnHisThrThrGlyValAsnCysGluArg-----CysLeu	409
QY	914	AAGAATTACTGCGAGTCTATGAGGCCCAATATGCTGCTCTTCTTCAATGCAAT	973
Db	410	ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgCys	429
QY	974	GGTTGCAAAAT-----TATGAAGAAGCCAGACGAAAGACACTA	1015
Db	430	AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys	447
QY	1016	ATGAGCATGCCAACTAC-----ATGCAGACTGGAGTTTGAAGGC	1057
Db	448	TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly	467
QY	1058	AGCCATTACCTGCCACCAACGAAATTTTCAGGACTTCCAGAGATTCAGTCAGTAGCGG	1117
Db	468	PheProSerCysTyrProThr-----ProSerSerAsnAspThrArg	482
QY	1118	CCTTCTCATCATCTCCTGGGAGGTGGTGGAGGCCACATCGCGCTGCCTGCTGCTCAG	1177

```
Db 483 GluGlnValLeuProAlaGlyGlnIleValAsnCysSerAlaAlaGlyThrGln 502
QY 1178 GGAGAAGAGCCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGATCCTGGAG 1237
Db 503 GlyAsnAlaCysArgLysAsp-----Pro 510
QY 1238 GAATTTGGAAGTGCTTTATCA-----CAGATTCTCCACACTGAGTTTAAATCT 1285
Db 511 ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla 530
QY 1286 AAGGATTGAAAATGGAGTAGACTATATAAGTGTGAATGC 1324
Db 531 ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys 543
```

Search completed: April 21, 2003, 11:55:41
Job time : 139.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:51:54 ; Search time 24.5 Seconds
(without alignments)
5125.593 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcggggtcaagcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlip
-Q=/cgn2_1/USPTO_spoos/US09743237/runat_21042003_112211_1508/app_query.fasta_1.2311
-DB=Issued_Patents_AA -QFWT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=PTO -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 -CGN 1.1.33 -runat_21042003_112211_1508 -NCRU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	124	3.2	739	4	US-08-444-818-148
2	124	3.2	2995	4	US-08-444-818-138
3	117	3.1	1021	1	US-07-910-760-12
4	117	3.1	1021	1	US-08-440-519-12
5	117	3.1	1021	4	US-08-440-549-12
6	117	3.1	2261	4	US-08-444-818-66
7	117	3.1	2436	4	US-08-444-818-75
8	117	3.1	2772	4	US-08-444-818-89
9	117	3.1	2894	2	US-08-466-975A-23
10	117	3.1	2894	2	US-08-391-671A-23
11	117	3.1	2894	3	US-08-467-902A-23
12	117	3.1	2894	4	US-09-275-265-23

13	117	3.1	3011	1	US-08-440-103-36	Sequence 36, Appl
14	117	3.1	3011	1	US-08-440-542-36	Sequence 36, Appl
15	117	3.1	3011	1	US-07-910-760-10	Sequence 10, Appl
16	117	3.1	3011	1	US-08-440-519-10	Sequence 10, Appl
17	117	3.1	3011	1	US-08-231-368-36	Sequence 36, Appl
18	117	3.1	3011	1	US-08-440-210-36	Sequence 36, Appl
19	117	3.1	3011	4	US-09-388-874-2	Sequence 2, Appl
20	117	3.1	3011	4	US-09-046-604-36	Sequence 36, Appl
21	117	3.1	3011	4	US-08-440-549-10	Sequence 10, Appl
22	117	3.1	3011	4	US-08-850-328-1	Sequence 1, Appl
23	116	3.0	1692	4	US-09-263-933-4	Sequence 4, Appl
24	116	3.0	1692	4	US-09-263-933-11	Sequence 11, Appl
25	116	3.0	1692	4	US-09-263-933-18	Sequence 18, Appl
26	116	3.0	2013	1	US-08-324-977-12	Sequence 12, Appl
27	116	3.0	2013	2	US-08-384-616-12	Sequence 12, Appl
28	116	3.0	2013	2	US-08-904-686A-12	Sequence 12, Appl
29	116	3.0	2013	4	US-09-315-850-12	Sequence 12, Appl
30	116	3.0	2201	4	US-08-952-981A-2	Sequence 2, Appl
31	116	3.0	2307	4	US-09-263-933-2	Sequence 2, Appl
32	116	3.0	2307	4	US-09-263-933-9	Sequence 9, Appl
33	116	3.0	2307	4	US-09-263-933-16	Sequence 16, Appl
34	116	3.0	2620	1	US-08-324-977-32	Sequence 32, Appl
35	116	3.0	2620	2	US-08-384-616-32	Sequence 32, Appl
36	116	3.0	2620	4	US-08-904-686A-32	Sequence 32, Appl
37	116	3.0	2620	4	US-09-315-850-32	Sequence 32, Appl
38	116	3.0	2621	1	US-08-324-977-36	Sequence 36, Appl
39	116	3.0	2621	2	US-08-384-616-36	Sequence 36, Appl
40	116	3.0	2621	2	US-08-904-686A-36	Sequence 36, Appl
41	116	3.0	2621	4	US-09-315-850-36	Sequence 36, Appl
42	116	3.0	3010	1	US-08-324-977-2	Sequence 2, Appl
43	116	3.0	3010	1	US-08-324-977-14	Sequence 14, Appl
44	116	3.0	3010	2	US-08-384-616-2	Sequence 2, Appl
45	116	3.0	3010	2	US-08-384-616-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-444-818-148
; Sequence 148, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 148:

SEQUENCE CHARACTERISTICS:
 LENGTH: 739 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-444-818-148

Alignment Scores:
 Pred. No.: 0.00155 Length: 739
 Score: 124.00 Matches: 91
 Percent Similarity: 35.32% Conservative: 45
 Best Local Similarity: 23.64% Mismatches: 140
 Query Match: 3.24% Indels: 109
 DB: 4 Gaps: 22

US-09-743-237-3 (1-2134) x US-08-444-818-148 (1-739)

```

QY 641 TCTGTTTCCCTCCGATCACTCTCCAGGACCCACCAAAATAACTTTGGCTGGG 697
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 233 SerValIleProThrSerGlyAspValValValValValAlaThrAspAlaLeuMetThrGly 252
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 698 TACTGTCACGCTTGGCCACTGGGACTTTTGCACAACTGCACATTTGTAATAATGTTGC 757
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 267
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 AACAACTTG-----CATCATGATATTGACGGTTTAAAGCCATT 796
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 287
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 797 AAGGCATGCTTGTGTAAGATCCAGAGCTTCCAGCCAAAATTTGGGAAGGCCCAATTG 856
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThr 300
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 857 GGCATGTCAGCCAGCCAGCACACAAA-----GGGTGCAACTGCAGAGGTGAGGCTGC 910
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GlyArgGlyLysProGlyIleAsnArgPheValAlaProGlyGluArgProSerGlyMet 320
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 911 CTGAAGAAT-----TACTCGAGTCTATGAGGCCCAATTTATGTTCTTCTATTGTC 964
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 321 PheAspSerValLeuLysGlyCysGlyCysTyrAspAla----- 332
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 965 AAATGCATTTGTTGCAAAATTTAAGAAAGCCAGAGCAAGAACACACTAATGAGCATG 1024
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 -----GlyCysAlaTrpTyrGluLeuThrPro--AlaGluThrThrValArgLeu 348
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1025 CCAACTACATGACAGCTGAGGTTTG-----GAA 1054
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 ArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGlu 368
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1055 GGC-----AGCCATTACTGCCCAACCAAGCAATTTTCA 1087
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 369 GlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSer 388
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1088 GGA-----CTTCCAAGATTGATC-----ACGATAGGGCGCTTCTCTCATGC 1129
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 389 GlyGluAsnLeuProThrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAl 408
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1130 ATCTCTGGGAGTGG-----TGGAGGCCACATGCGCTGCTGCTGCTGCTGCTGCTGCTG 1180
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 aProProSerTrpAspGlnMetTrpLys-----CysLeuIleArgLe 423
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1181 GAAGAGGCCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGACAGATGATCCTGGAGGAA 1240
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 uLysProThrLeuHisGlyProThrProLeuLeuThrArgLeu----- 437
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1241 TTGGAGGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 438 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMe 455
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1278 -----TTAAATCTAGGAGTGAATAATGGAGTAGAGTATAAGTGT 1318
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 tThrCysMetSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyGlyVa 475
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 1319 GAATGCA-----TGTTGATTTTCTCTAGTAATCTCTAGTTTAGAAGATGTT 1372
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 lLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 492
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1373 TAGGGGAACATGAGCTGGCTCTGCAGCAACAACACAGGCTCCCTGCATCCCTGGGCCCA 1432
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 493 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArgG 508
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1433 GGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTCTGA 1486
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 uValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIleG 528
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1487 GGAGTGCATGGCTGAGCATTTGTTCTCTGGCCCAAGAGAGAGAGCTGGGTTCACATA 1546
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 uGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu----- 544
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1547 GTCTGGGAGAGTGTCTGCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCACT 1606
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 545 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnThrAs 562
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1607 CTGCTGCACTCTT 1619
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 562 nTrpGlnLysLeu 566
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 2

```

US-08-444-818-138
; Sequence 138, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2995 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-138

```

Alignment Scores:
 Pred. No.: 0.00286 Length: 2995
 Score: 124.00 Matches: 91
 Percent Similarity: 35.32% Conservative: 45
 Best Local Similarity: 23.64% Mismatches: 140
 Query Match: 3.24% Indels: 109

```
DB: 4 Gaps: 22
US-09-743-237-3 (1-2134) x US-08-444-818-138 (1-2995)
QY 641 TCTGTTTCCCC---TCGGATCAACTCTTCAGGACACCAAAATAACTTTGGCTGG 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1424 SerValIleProThrSerGlyAspValValValAlaThraAspAlaLeuMetThrGly 1443
QY 698 TACTGTGACTGTTGCCAGTGGGACATTTTCACAACTGCAATTAATAATTGTTGC 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
DB 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
QY 797 AAGCATGTCTGGTAGAATCCAGAGCTTTCAGCCCAAAATTTGGGAAGGCCAAATG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThr 1491
QY 857 GCAATGTCAAGCCCGACACACAAA-----GGGTGCACTGCAGGAGGTTCAGCTGC 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1492 GlyArgGlyProGlyIleAsnArgPheValAlaProGlyGluArgProSerGlyMet 1511
QY 911 CTGAAGAAT-----TACTGCGAGTCTATGAGGCCAAATTTATGTTCTTCTATTGCG 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1512 PheAspSerValLeuCysGluCysArgThrAspAla----- 1523
QY 965 AATGCTATTGTTGAAAATTTATGAAGAACCCAGAACGACACACTAATGACATG 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1524 -----GlyCysAlaTrpArgGluLeuThrPro-----AlaGluThrThrValArgLeu 1539
QY 1025 CCAAACTACATCAGACGTGGAGTTG-----GAA 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1540 ArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGlu 1559
QY 1055 GGC-----AGCCATTACTGCCACCAACGAAATTTTCA 1087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1560 GlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSer 1579
QY 1088 GGA-----CTTCCAAGATTACGTC-----ACGATAGGCGGCTTCCCTCATGC 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1580 GlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAl 1599
QY 1130 ATCTCTGGGAGGTGG-----TGGAGGCCACATCGCGCTGCTGCTCTCAGGGA 1180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1599 aProProSerTrpAspGlnMetTrpLys-----CysLeuIleArgLeu 1614
QY 1181 GAAGAGCGCGAGAAACACTGCTCCAAGTGCCTGGCAGACAGATGATCCTGGAGGAA 1240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1614 uLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
QY 1241 TTTGAAGTGTATTATCAGATTCTCCACACTGAGT----- 1277
DB 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIleMe 1646
QY 1278 -----TTAAATCTAAGGATTGAAATGAGTACGATGATTAAGTCT 1318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1646 tThrCysMetSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyGlyVa 1666
QY 1319 GAATGCA-----TGTGATTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAGGATGTT 1372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1666 lLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
QY 1373 TAGGGAACATGAGCTGCTGTCAGCAACCAACAGGCTCCCTGCAATCCCTGGGCCA 1432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArgG 1699
QY 1433 GGGAGTTTACTCAGAGCTCTCT-----CAAGATGTGGCAACCCATGCCCTTTCTGCA 1486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1699 uValLeuTyrArgGluPheAspGluMetGluCysSerGlnHisLeuProTyrIleG 1719
QY 1487 GGAGGTGCATGCGCTGAGCATTTGTTGCTGGCCCGAGAGGAGAGCTTGGTTCCTCAT 1546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 446
QY 797 AAGGATGCTTGGTAGAATCCAGAGCTTCCAGCCAAAATTTGGGAAGGCCCAATTG 856
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAGCCCGCAGCACACAAAGGTCGAAGTGGAGGAGGTCAGGC 907
Db 465 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAATATGTTCTTCTATT 961
Db 479 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 491
QY 962 TCAAAATGATTTGGTGAATAATATGAAGAAAGCCAGAACGAAACACACATATGAGC 1021
Db 492 -----GlyCysAlaTyrPyrGluLeuThrPro---AlaGluThrThrValArg 506
QY 1022 ATGCCAAACTACATCAGCTGAGGAGTTG----- 1051
Db 507 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
QY 1052 GAAGGC-----AGCCATTACCTGCCACCAACGAAATTT 1084
Db 527 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 546
QY 1085 TCAGGA-----CTTCAAGATTCAGTC-----ACGATAGCGGCCCTTCCTCA 1126
Db 547 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 566
QY 1127 TGCATCTCTGGGAGGTG-----TGGAGCCACATCGGCTGCTGCTGCTGCTCAG 1177
Db 566 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg 581
QY 1178 GGAGAAGAGCGGAGAAAGACACTGCTCCAAGTGCCTGGCAGCAGCATGATCCTGGAG 1237
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 596
QY 1238 GAATTTGAAGGTGCTTATCAGATTCACAGATTCACACTGAGT----- 1277
Db 597 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 613
QY 1278 -----TTAAATCTAAGGATTAAGATTAAGATTAAGATTAAGATTAAG 1315
Db 613 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyL1 633
QY 1316 TGCAATGCA-----TGTTGATTTGCTTAGTCTAGAAATCTCTAGTTTGAAGAGGAT 1369
Db 633 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 651
QY 1370 GTTTAGGGACATGAGCTGGCTCTGCAGCAACACAGCTCCCTGCTGATCCCTGGC 1429
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 666
QY 1430 CCAGGGAGTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC 1483
Db 666 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 686
QY 1484 TGAGAGGTGCATGCTGAGCATTTGTTGCTGGCCAGGAGGAGAGAGCTTGGGTTCC 1543
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 703
QY 1544 ATAGTCTGGAGAGTGTCTGCAGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1603
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla----ProAlaValGlnTh 720
QY 1604 ACTGCTGCTGACTCTT 1619
Db 720 rAsnTrpGlnLysLeu 725

```

RESULT 4

US-08-440-519-12

; Sequence 12, Application US/08440519

```

; Patent No. 5712087
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,519
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/910,760
; FILING DATE: 07-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1021 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-519-12

```

```

Alignment Scores:
Pred. No.: 0.00863
Score: 117.00
Percent Similarity: 35.49%
Best Local Similarity: 23.58%
Query Match: 3.06%
DB: 111
Gaps: 23

```

US-09-743-237-3 (1-2134) x US-08-440-519-12 (1-1021)

```

QY 641 TCTGCTTCCCC---TCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 392 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 411
QY 698 TACTGTGACTGCTTCCGACTGGGACTTTTGCACCAACTGCAATTTGTAATAATTGTTGC 757
Db 412 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 426
QY 758 AACCAACTG-----CATCATGATTTCAACGGTTTAAGCCATT 796
Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 446
QY 797 AAGCATGTCTTGGTGAATAATCCAGAGCTTTCAGCCAAAATTTGGGAAGGCCCAATTG 856
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAGGCCCCAGCAGCACACAAAGGTCGAAGTGGAGGAGGTCAGGC 907
Db 465 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTCTGAGGCCCAATATGTTCTTCTATT 961

```



```
Db 479 MetPheAspSerSerValLeuCysGluCysTyAspAla----- 491
QY 962 TGCAAAATGCTGGTTGCAAAATATGAAGAAACCCAGACGAAAGACACTAATGAGC 1021
Db 492 -----GlyCysAlaTrpTrpGluLeuThrPro---AlaGluThrThrValArg 506
QY 1022 ATGCCAAATCACTACAGACTGAGGTTTG----- 1051
Db 507 LeuArgAlaTyMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
QY 1052 GAAGGC----- 526
Db 527 GluGlyValPheThrGlyLeuThrHisLeuAspAlaHisPheLeuSerGlnThrLysGln 546
QY 1085 TCAGCA-----CTTCCAGATTCAGTC-----AGCATAGGGGGCTTCTCTCA 1126
Db 547 SerGlyGluAsnLeuProTyLeu-ValAlaTyTrpGlnAlaThrValCysAlaArgAlaG1 566
QY 1127 TGCAATCTCTGGAGGTGG-----TGGAGGCCACATGGCGCTGCTGCTGCTCTCAG 1177
Db 566 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg 581
QY 1178 GGAGAGAGCGCGAAGAACACTGCTCCAGTCCCTGGCAGACAGATGATCCTGGAG 1237
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyArgLeu----- 596
QY 1238 GAATTTGGAAGTGCTTATCACAGATTCCTCCACACTGAGT----- 1277
Db 597 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyTr1 613
QY 1278 -----TTAAATCTAAGGATTGAAAATGGAGTAGAGTATAAAG 1315
Db 613 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGly1 633
QY 1316 TGTGAATGCA-----TGTGATTTGCTTAGTCTAGTAATCTCTAGTTTGAAGAGAT 1369
Db 633 yValLeuAlaAlaLeuAlaAlaTyCysLeuSerThrGlyCysValValIleVal----- 651
QY 1370 GTTTAGGGGAACATGAGCTGCTGCGAGCAACACAGGCTCCCTGCATCCCTGGGC 1429
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleProAspAr 666
QY 1430 CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC 1483
Db 666 gLluValLeuTyArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyTr1 686
QY 1484 TGAGGAGTGCATGCCCTGAGCAATTTGTTGCTGCGCCAGAGGAGAGCTTTGGGTTCCC 1543
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 703
QY 1544 ATAGTCCTGGGAGAGTGTGCGAGGCGCGGAGGCGAGCAGCAGGCTCGCGAGAGCTC 1603
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 720
QY 1604 ACTCTGGTCCGACTCTT 1619
Db 720 rAsnTrpGlnLysLeu 725

RESULT 5
US-08-440-549-12
; Sequence 12, Application US/08440549
; Patent No. 6312889
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
```

```
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,549
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/910,760
; FILING DATE: 07-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1021 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-549-12

Alignment Scores:
Pred. No.: 0.00863 Length: 1021
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 4 Gaps: 23
```

US-09-743-237-3 (1-2134) x US-08-440-549-12 (1-1021)

```
QY 641 TCTGCTTTCCCC---TCGGGATCAACTCTTCCAGGACCACCAAAATAAATTTGGCTGGG 697
Db 392 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 411
QY 698 TACTGTGACTGTTTCCAGTGGGACTTTTCCACAACTGCAATTTGAATTAATTTGTTGC 757
Db 412 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 426
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 427 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 446
QY 797 AAGGCATCTCTGGTAGAAATCCAGAAAGCTTTCCAGCCAAATAATTTGGGAAGGCCCAATTG 856
Db 447 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 464
QY 857 GGC-----AATGTCAGCCGCCAGCACAAAGGTCGCACTGCAGGAGGTGAGGC 907
Db 465 GlyIleTyArgPheValAlaProGlyGlu-----ArgProSerGly 478
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTATGAGGCCCAATTTATGTGTTCTTCTATT 961
Db 479 MetPheAspSerSerValLeuCysGluCysTyAspAla----- 491
QY 962 TGCAAAATGCTGGTTGCAAAATATGAAGAAACCCAGACGAAAGACACTAATGAGC 1021
Db 492 -----GlyCysAlaTrpTrpGluLeuThrPro---AlaGluThrThrValArg 506
QY 1022 ATGCCAAATCACTACAGACTGAGGTTTG----- 1051
Db 507 LeuArgAlaTyMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 526
```

QY 1052 GAAGGC-----AGCATTACCTGCCCAACGAAATTT 1084
||||| :|||:|||||
Db 527 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 546
||||| :|||:|||||
QY 1085 TCAGGA-----CTTCCAAGATTGAGTCT 1126
||||| :|||:|||||
Db 547 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 566
||||| :|||:|||||
QY 1127 TGCATCTCCGGGAGTGG-----TGGAGCCACATCGCGCTCCCTGCTGCTCAG 1177
||| ||| ||| :|||:|||||
Db 566 nAlaProProSerThrAspGlnMetTrpLys-----CysLeuIleAr 581
||||| :|||:|||||
QY 1178 GGAGAAGAGCCGCGAAGAAGCACTGCTCCAAAGTCCCTGGCAGCAGATCATCTCGGAG 1237
||| ||| ||| :|||:|||||
Db 581 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 596
||||| :|||:|||||
QY 1238 GAATTTGGAGGTCTTATCATCAGATTCTCCACACTAGT----- 1277
||||| :|||:|||||
Db 597 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 613
||||| :|||:|||||
QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGAGTATAAAG 1315
||| :|||:|||||
Db 613 eMetThrCysMetSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyG1 633
||||| :|||:|||||
QY 1316 TGTGAATGCA-----TGTGATTCTCTAGTCTAGATAATCTCTAGTTTAGAAAGGAT 1369
||| ||| ||| :|||:|||||
Db 633 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal----- 651
||||| :|||:|||||
QY 1370 GTTTAGGGGAACATGAGCGTCTCCAGCAACACGAGTCCCTGCTGATCCCTGGGC 1429
||| :|||:|||||
Db 652 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 666
||||| :|||:|||||
QY 1430 CCAGGAGTTTACTCAGACTCTCT-----GAGATGTGSCAACCCATGCCCTTTTC 1483
||| :|||:|||||
Db 666 gLeuValLeuTyrArgGluPheAspGluMetGluGlyCysSerGlnHisLeuProTyrI1 686
||||| :|||:|||||
QY 1484 TGAGGAGTGATGCGCTGAGCAATGTTGTGTGCGCCAGAGGAGAGAGCTTGGGTTCCC 1543
||| :|||:|||||
Db 686 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 703
||||| :|||:|||||
QY 1544 ATAGTCTGGGAGAGTCTGCGAGGCGCGGAGGAGCAGCAGCGCTCGGAGAGCTC 1603
||| :|||:|||||
Db 704 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 720
||||| :|||:|||||
QY 1604 ACTGTGTGCACTCTT 1619
||| ||| |||
Db 720 rAsnTrpGlnLysLeu 725
||||| :|||:|||||

RESULT 6

US-08-444-818-66
; Sequence 66, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2261 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-66

Alignment Scores: Pred. No.: 0.0122 Length: 2261
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 4 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-444-818-66 (1-2261)
QY 641 TCTGCTTTCCCC---TCGGGATCACTCTTCAGGACCACCAAAATAACTTTGGCTGGG 697
||| ||| ||| :|||:|||||
Db 799 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 818
QY 698 TACTGTGACGTCTGCCAGTGGGACTTTTGCACAACTGCAATTCGTAATAATTTGTCG 757
||| :|||:|||||
Db 819 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 833
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
||| :|||:|||||
Db 834 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 853
QY 797 AAGCATGCTTGGTAGAAATCCAGAGCTTTCAGCCAAATAATGGAGAGGCCAATTG 856
||| :|||:|||||
Db 854 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 871
QY 857 GGC-----AATGTCAAGCCCGACAGCAACAAGGGTGCAACTCGACGAGTGCAGGC 907
||| ||| ||| :|||:|||||
Db 872 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 885
QY 908 TGCCTGAAGAAT-----TACTCGAGTGTATGAGGCCCAAAATATGTCTTCTTATT 961
||| :|||:|||||
Db 886 MetPheAspSerValLeuCysGluCysTyrAspAla----- 898
QY 962 TGCAAATGCAATGCTTGCACAAATAATATGAAGAACCCCAACAGACACACATGAGC 1021
||| :|||:|||||
Db 899 -----GlyCysAlaTyrTyrGluLeuThrPro---AlaGluThrThrValArg 913
QY 1022 ATGCCAAACTACATCAGACTGCGAGTGTG----- 1051
||| :|||:|||||
Db 914 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 933
QY 1052 GAAGGC-----AGCCATTACCTGCCCAACGAAATTT 1084
||||| :|||:|||||
Db 934 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 953
QY 1085 TCAGGA-----CTTCCAAGATTGAGTCT-----ACGATAGCGCGCTTCTCTCA 1126
||||| :|||:|||||
Db 954 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 973
QY 1127 TGCATCTCCCTGGGAGTGG-----TGGAGGCCACATCGCGCTCCCTGCTGCTCAG 1177
||| ||| ||| :|||:|||||
Db 973 nAlaProProSerThrAspGlnMetTrpLys-----CysLeuIleAr 988
QY 1178 GGAGAAGAGCCGCGAAGAAGCACTGCTCCAAAGTGTGCGAGAGCAGATGATCTCGGAG 1237

```

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-75

Alignment Scores:
Pred. No.: 0.0127 Length: 2436
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 4 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-444-818-75 (1-2436)
QY 641 TCTGCTTTCCCC--TCGGGATCACTCTTCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 974 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 993
QY 698 TACTGTGACTGCTTTGCCAGTGGGACTTTTGCACCAACTGCAATTTGTAATATTTGTC 757
Db 994 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1008
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 1009 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1028
QY 797 AAGGATGCTCTTGGTAGAAATCCAGNAGCTTTCCAGCCCAAAATTTGGGAAGGCCAATTG 856
Db 1029 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1046
QY 857 GGC-----AATGTCAGGCCCCAGCAGCACAAAGGGTGCAACTGCAGGAGGTCCAGGC 907
Db 1047 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1060
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGCTATGAGGCCCAATAATATGTCTTCTATT 961
Db 1061 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1073
QY 962 TGCAAATGCTGTTGGCAAAATTTATGAAGAAAGCCAGAACGAAGACACATAATGAGC 1021
Db 1074 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg 1088
QY 1022 ATGCCAACTACATGCAGACTGGAGGTTG----- 1051
Db 1089 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1108
QY 1052 GAAGC-----AGCCATTACCTGCCACCAAGCAAAATTT 1084
Db 1109 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1128
QY 1085 TCAGGA-----CTTCCAAGATTTCAGTC-----ACGATAGGCGGCTTCTCTCA 1126
Db 1129 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG 1148
QY 1127 TCATCTCTCTGGGAGGTGG-----TGGAGGCCACATGCGCCTGCTGTGCTCAG 1177
Db 1148 nAlaProProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1163
QY 1178 GGAGAAGAGGCCGAGAAAGAACACTCTCTCCAAAGTGCTGCGCAGAGCAGATGATCTGGAG 1237
Db 1163 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1178
QY 1238 GAATTTGGGAAGGTGCTTATCAGAGATTCACACAGT----- 1277
Db 1179 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI 1195
QY 1278 -----TTAAATCTAAGGGATTGAAATGCGAGTAGATATAAAG 1315
Db 1195 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG 1215
QY 1316 TGTGAATGCA-----TGTGATTGTTGCTTAGCTAGAAATCTCTAGTTTAAAGAGAT 1369
;
; LENGTH: 2436 amino acids
;
```

RESULT 7

```

US-08-444-818-75
; Sequence 75, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2436 amino acids
;
```

Db 1215 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1233
Qy 1370 GTTTAGGGGAACATGAGCTGGCTCTGCAGCAACAACAGGCTCCCTGCATCCCTGGG 1429
Db 1234 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1248
Qy 1430 CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGCAACCCATGCTCCCTTTTC 1483
Db 1248 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI 1268
Qy 1484 TGAGGAGTGCATGCCCTGACCATTTTGTCTGCCCGAGGAGAGAGCTTGGGTCCC 1543
Db 1268 eGluGln-----GlyMetLeuAlaGluGlnPheLysGlnHisLeuGlyLeu-- 1285
Qy 1544 ATAGTCTGGGAGAGTCTCCAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1603
Db 1286 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1302
Qy 1604 ACTCTGGTCTGCTT 1619
Db 1302 rAsnTrpGlnLysLeu 1307

RESULT 8

US-08-444-818-89
; Sequence 89, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David V.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-89

Alignment Scores:
Pred. No.: 0.0134 Length: 2772
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 4 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-444-818-89 (1-2772)
Qy 641 TCTGCTTTCCCTCCGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 1310 SerValIleProThrSerGlyAspValValValValAlaThrAspAlaLeuMetThrGly 1329
Qy 698 TACTGTGACTGCTTGGCAGTGGGACTTTTGCACCAACTGCAATTTGTAATAATTGTTGC 757
Db 1330 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1344
Qy 758 AACAACTTG-----CATCATGATATTGACCGGTTTAAGCCATT 796
Db 1345 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1364
Qy 797 AAGCATGTCTTGTAGAAATCCAGAAGCTTCCAGCCAAAATTTGGGAAGGCCAATTG 856
Db 1365 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1382
Qy 857 GGC-----AATGTCAAGCCCCCAGCAGCAACAAGGGTGCACTGCAGGAGGTCAAGC 907
Db 1383 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1396
Qy 908 TGCCTGAAGAAT-----TACTGGAGTCTGTATGAGCCCAAAATATTGCTTCTTCTATT 961
Db 1397 MetPheAspSerValLeuCysGluCysTyrAspAla----- 1409
Qy 962 TGCAAAATGCATTGTTGCAAAAATTATGAAGAAGCCAGAACAGAAAGACACATAATGAGC 1021
Db 1410 -----GlyCysAlaTyrTrpGluLeuThrPro--AlaGluThrThrValArg 1424
Qy 1022 ATGCCAACTACATGCAGACTGGAGGTTTG----- 1051
Db 1425 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1444
Qy 1052 GAAGGC-----AGCATTAACCTGCACCAACCAAGAAATTT 1084
Db 1445 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1464
Qy 1085 TCAGGA-----CTTCCAAGATTACAGTC-----ACGATAGCGCGCTTCCTCA 1126
Db 1465 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG 1484
Qy 1127 TGCATCTCTCTGGGAGGTG-----TGGAGGCCACATCGCGCTCCCTGCTGCTGCTCAG 1177
Db 1484 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1499
Qy 1178 GGAGAGAGCGCGAAGAAAGAACACTGCTCCAAAGTCTGCGAGCAGAGAGATGATCCTGGAG 1237
Db 1499 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1514
Qy 1238 GAATTTGGAAGGTCTTATCACAGATTCTCCACACTGAGT----- 1277
Db 1515 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI 1531
Qy 1278 -----TTAAATCTAAGGATTGAAAATGGAGTAGAGATATAAG 1315
Db 1531 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGly 1551
Qy 1316 TGTGAATGCA-----TGTTGATTTTCTTAGTCTAGAAATCTCTAGTTTAGAAGGAT 1369
Db 1551 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1569
Qy 1370 GTTTAGGGGAACATGAGCTGGCTCTGCAGCAACAACAGGCTCCCTGCTGCTGCTGGC 1429
Db 1570 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1584
Qy 1430 CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGCAACCCATGCTCCCTTTTC 1483
Db 1584 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI 1604
Qy 1484 TGAGGAGTGCATGCCCTGACCATTTTGTCTGCCCGAGGAGGAGGAGGAGGAGGAGGAG 1543
Db 1604 eGluGln-----GlyMetLeuAlaGluGlnPheLysGlnHisLeuGlyLeu-- 1621

QY 1544 ATAGTCTGGAGAGTGTCTGCAGGGGGGAGGAGCAGCAGCCCTGCGGAGAGCTC 1603
Db 1622 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1638
QY 1604 ACTCTGGTGCAGCTCTT 1619
Db 1638 rAsnTrpGlnLysLeu 1643

RESULT 9

US-08-466-975A-23
; Sequence 23, Application US/08466975A
; Patent No. 5910404
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,975A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,671
; FILING DATE:
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEtical: NO
; ANTI-SENSE: NO
US-08-466-975A-23

Alignment Scores:
Pred. No.: 0.0137 Length: 2894
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 45
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-466-975A-23 (1-2894)
QY 641 TCTGCTTTCCCC---TCGGGATCAACTTCTTCAGGACCACCAAAAAATACTTTGGTGGG 697
Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
QY 698 TACTGTGACTGCTTTCCCGAGTGGGACTTTTGCACAACTGCAATTCGTAATATCTTGC 757
Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
QY 758 AACAACTTG-----CATCATGATATTCAAGCGTTTAAAGCCATT 796
Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrLeuPro 1478
QY 797 AAGGCATGCTTGGTAGAAATCCAGAGCTTTCCAGCCAAAATTTGGGAAGGCCCAATTG 856
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
QY 857 GGC-----AATGTCAGGCCCCAGCAGCACAAAGGGTGCACATGCAGGAGTCCAGC 907
Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
QY 908 TGCCTGAAGAAT-----TACTGCGAGTGTCTATGAGCCCAAAATATGCTTCTTCTATT 961
Db 1511 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1523
QY 962 TGCAAATGCTTGGTTGCAAAAATTTATGAAGAAAGCCAGACAGAAAGACATAATGAGC 1021
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg 1538
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTG----- 1051
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
QY 1052 GAAGGC-----AGCCATTACTGCGCACCAACAGCAAAATTT 1084
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGly 1578
QY 1085 TCAGGA-----CTTCCAAGATTTCAGTC-----ACGATAGGGGGCCCTTCCCTCA 1126
Db 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaAlaArgAla 1598
QY 1127 TGCATCTCTCTGGGAGGTGG-----TGGAGGCCACATGCGCCTGCTGCTGCTGCTCAG 1177
Db 1598 nAlaProProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1613
QY 1178 GGAGAAGAGCCCGAGAAAGAACACTGCTCCAGTGCCTGCGCAGACAGATGATCTGGAG 1237
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
QY 1238 GAATTTGGGAGGTGCTTATCAGAGATTCTCCACACTGAGT----- 1277
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle 1645
QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGAGATATAAAG 1315
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG 1665
QY 1316 TGTGAATGCA-----TGTGTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAAGAT 1369
Db 1665 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal----- 1683
QY 1370 GTTTAGGGGAACATGAGGCTGGCTCTGCAGCACACACAGGCTCCCTGCATCCCTGGGC 1429
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
QY 1430 CCAGGAGGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCCCTTTTC 1483
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIle 1718
QY 1484 TGAGGAGGTGCATGGCCTGAGCATTTGTTTCTGGCCCCCAGGAGAGAGCTGGGTGCC 1543
Db 1484 ----- 1543

Db 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
Qy 1544 ATAGTCTGGAGAGTCTCTCAGGGCGGCGAGGCGAGAGCGGCTCGGGAGAGCTC 1603
Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
Qy 1604 ACTGTGGTGCAGCTTT 1619
Db 1752 rAsnTrpGlnLysLeu 1757

RESULT 10

US-08-391-671A-23
; Sequence 23, Application US/08391671A
; Patent No. 5922532
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWON, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHIE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,671A
; FILING DATE: 21-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:

; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

; LENGTH: 2894 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-391-671A-23

Alignment Scores:

Pred. No.: 0.0137 Length: 2894
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-391-671A-23 (1-2894)
Qy 641 TCTGCTTTCCCC---TCGGGATCAACTCTTCCAGGACCACAAATAAATACTTTGGCTGGG 697
Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
Qy 698 TACTGTGACTGCTTTCAGTGGGACTTTTGCACAACTGCAATTTGTAATAATTTGTTGC 757
Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
Qy 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
Qy 797 AAGCATGTCTTGGTAGAAATCCAGAAGCTTTCAGGCAAAATTTGGGAAGGGCAATTG 856
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyPro 1496
Qy 857 GGC-----AATGTCAAGCCCGACGACACAAAGGGTGCACTCGACGAGGTTCAGGC 907
Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
Qy 908 TGCCTGAAGAAT-----TACTGGGAGTGTATGAGCCCAAAATATTGTCTTCTATT 961
Db 1511 MetPheAspSerValLeuCysGluCysTyrAspAla----- 1523
Qy 962 TGCAAAATGCATTGTGTCAGAAATATTATGAAGAACCCAGAACAGACACTAATGAGC 1021
Db 1524 -----GlyCysAlaTyrTyrGluLeuThrPro--AlaGluThrThrValArg 1538
Qy 1022 ATGCCAACTACATGCAGACTGGAGTTTG----- 1051
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
Qy 1052 GAAGGC-----AGCCATTACCTGCCACCACCAACGAAATTT 1084
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGly 1578
Qy 1085 TCAGGA-----CTTCCAAGATTACAGTC-----ACGATAGCGGCGCTTCCTCA 1126
Db 1579 SerGlyGlnAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG 1598
Qy 1127 TGCATCTCTCTGGGAGTGT-----TGGAGGCCACATCGCGCTCGCTTGTCTGCAG 1177
Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1613
Qy 1178 GGAGAGAGCCCGAGAGAACACTGCTCCAAGTCTGCGAGAGCAGATGATCCTGGAG 1237
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
Qy 1238 GAATTTGGAAGTGTCTTATCAGAGATTCTCCACACTCAGT----- 1277
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI 1645
Qy 1278 -----TTAAATCTAAGGATTGAAAATGGAGTAGAGTATAAAG 1315
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGly 1665
Qy 1316 TGTCAATGCA-----TGTTGATTTTGTCTAGTCTAGAATCTCTAGTTTAGAAAGAT 1369
Db 1665 yValLeuAlaAlaLeuAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
Qy 1370 GTTTAGGGGAACATGAGGCTGCTGCAGCAACACAGGCTCCCTGCATCCTCGGCG 1429
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
Qy 1430 CCAGGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCTTTTC 1483
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI 1718
Qy 1484 TGAGGAGTGCATGGGCTGAGCATTGTTTCTCGCCCGAGAGGAGAGCTTGGTTCCC 1543
Db 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735

Db 1718 eGluGln-----GlyMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
QY 1544 ATAGTCTGGAGAGTGTCTGCAGGGCGGGAGGCGGAGAGGCGCTCGGAGAGCTC 1603
Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
QY 1604 ACTCTGGTCGACTCTT 1619
Db 1752 rAsnTrpGlnLysLeu 1757

RESULT 13

US-08-440-103-36
; Sequence 36, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-36

Alignment Scores:
Pred. No.: 0.0139 Length: 3011
Score: 117.00 Matches: 91
Percent Similarity: 35.49% Conservative: 46
Best Local Similarity: 23.58% Mismatches: 138
Query Match: 3.06% Indels: 111
DB: 1 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-440-103-36 (1-3011)

QY 641 TCTGCTTTCCCC--TCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGG 697
Db 1424 SerValleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
QY 698 TACTGTGACTGCTTGGCCAGTGGGGACTTTTGCACCAACTGCAATTGTAATAATTGTTGC 757
||| :||||| :|||

Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
QY 758 AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT 796
Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
QY 797 AAGGCATGCTCTGGTAGAAATCCAGAAGCTTTCCAGCCAAAATTTGGGAAGGGCCCAATTG 856
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
QY 857 GGC-----AATGTCAAGCCCCAGCACACAACAAAGGTGCAACTGCAGGAGGTCCAGGC 907
Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
QY 908 TGCCCTGAAGAAT-----TACTCGGAGTGTATGAGCCCAAAATTTATGCTTCTTCTATT 961
Db 1511 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1523
QY 962 TGCAAAATGCATTGTTGCAAAAATTTATGAAGAAAGCCACACGAAAGACACTAATGAGC 1021
Db 1524 -----GlyCysAlaTyrTyrGluLeuThrPro---AlaGluThrThrValArg 1538
QY 1022 ATGCCAAACTACATGCAGACTGGAGTTTG----- 1051
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
QY 1052 GAAGGC-----AGCATTTACCTGCCACCAACAGAAATTT 1084
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1578
QY 1085 TCAGGA-----CTTCCAAGATTTCAGTC-----ACGATAGCGCGCTCTCTCA 1126
Db 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 1598
QY 1127 TGCATCTCTGGAGGTGG-----TGAGGCCACATGCGCTGCTGCTGTGCTCAG 1177
Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1613
QY 1178 GGAGAAGAGCCCGAGAAAGAACACTGCTCCAAGTGCCTGCCAGAGCAGATGATCCTGGAG 1237
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
QY 1238 GAATTTGGGAAGGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1645
QY 1278 -----TTAAATCTAAGGGATTGAAATGGAGTAGAGTATAAAG 1315
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 1665
QY 1316 TGTGAATGCA-----TGTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAGGAT 1369
Db 1665 yValLeuAlaAlaLeuAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
QY 1370 GTTTAGGGGAACATGAGCTGCTGCCACACAACACCGCTCCCTCGCATCCCTGGGC 1429
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
QY 1430 CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATCCCTTTTTC 1483
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1718
QY 1484 TGAGGAGGTGCATGGCTGAGCAATTTGTTCTGTGGCCAGGAGGAGAGCTGTGGTTTCCC 1543
Db 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
QY 1544 ATAGTCTGGAGAGTGTCTGCAGGGCGGGAGGCGGAGAGGCGCTCGGAGAGCTC 1603
Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
QY 1604 ACTCTGGTCGACTCTT 1619
Db 1752 rAsnTrpGlnLysLeu 1757

RESULT 14

US-08-440-542-36
 ; Sequence 36, Application US/08440542
 ; Patent No. 5670153
 ; GENERAL INFORMATION:
 ; APPLICANT: Weiner, Amy J.
 ; APPLICANT: Houghton, Michael
 ; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA: US/08/440,542

APPLICATION NUMBER: US/08/440,542
 FILING DATE: 12-MAY-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/231,368

FILING DATE:

APPLICATION NUMBER: US 07/759,575

FILING DATE: 13-SEP-1991

ATTORNEY/AGENT INFORMATION:

NAME: McClung, Barbara G.

REGISTRATION NUMBER: 33,113

REFERENCE/DOCKET NUMBER: 0205.001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 601-2708

TELEFAX: (510) 655-3542

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 3011 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-440-542-36

Alignment Scores:
 Pred. No.: 0.0139 Length: 3011
 Score: 117.00 Matches: 91
 Percent Similarity: 35.49% Conservative: 46
 Best Local Similarity: 23.58% Mismatches: 138
 Query Match: 3.06% Indels: 111
 DB: 1 Gaps: 23

US-09-743-237-3 (1-2134) x US-08-440-542-36 (1-3011)

Qy 641 TCTGTTTCCCC---TCGGGATCAACTCTCCAGGACCACCAAAATAAATTCGCTGGG 697
 Db 1424 SerValIleProThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGly 1443
 Qy 698 TACTGTGACTGCTTGGCAGTGGGACTTTTGCACAACTCCAATTTGTAATATTTGTC 757
 Db 1444 Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal 1458
 Qy 758 AACAACTG-----CATCATGATATTAAGCGTTTAAAGCCATT 796
 Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
 Qy 797 AAGGCATGCTTGTGGTAGAAATCCAGAACTTCCAGCCAAAATAATGGGAAGGCGCAATTG 856
 Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496

Qy 857 GGC-----AATGTCAAGCCCCAGCAGCACAAACAAAGGTGCAACTGCAGGAGGTCAAGC 907
 Db 1497 GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
 Qy 908 TGCCTGAAGAT-----TACTGCGAGTGTATGATGAGCCCAAAATATGTGTTCTTATT 961
 Db 1511 MetPheAspSerSerValLeuCysGluCysTyrAspAla----- 1523
 Qy 962 TGCAATGATTTGTTGTCACAAATATGTAAGAAAGCCAGACAAAGACACAACTAATGAGC 1021
 Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg 1538
 Qy 1022 ATGCCAACTACATGCAGACTGGAGTTTG----- 1051
 Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
 Qy 1052 GAAGGC-----AGCATTACCTGCCACCAACCAAAATTT 1084
 Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1578
 Qy 1085 TCAGGA-----CTTCAAGATTTCAGTC-----ACGATAGCGCGCTTCTCTCA 1126
 Db 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 1598
 Qy 1127 TGCATCTCCTGGGAGTGG-----TGGAGCCACATGCGCCTGCCTGCTGCTCAG 1177
 Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleAr 1613
 Qy 1178 GGAGAAGAGCGCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGATCCTGGAG 1237
 Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
 Qy 1238 GAATTTGGAAGGTGCTTATCACAGATTCTCCACACTGAGT----- 1277
 Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1645
 Qy 1278 -----TTAAATCTAAGGATTTGAAATGGAATGGAGTAGAGTATAAAG 1315
 Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGly1 1665
 Qy 1316 TGTGAATGCA-----TGTTGATTTGCTTAGTCTAGAAATCTAGTTTAGAAAGAT 1369
 Db 1665 yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
 Qy 1370 GTTTAGGGGAACATGAGGCTGCTGCAGCAACAAACAGCTCCCTGCATCCTCCCTGGGC 1429
 Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspAr 1698
 Qy 1430 CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCCCTTTTC 1493
 Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1718
 Qy 1484 TGAGGAGTGCATGCGCTGAGCATTTGTTGTCGCGCCAGAGGAGAGAGCTTGGTGTCCC 1543
 Db 1718 eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-- 1735
 Qy 1544 ATACTCTGGGAGAGTGTCTGCAGGCGCGGAGGCGAGAGCGCGCTGCGAGAGAGCTC 1603
 Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
 Qy 1604 ACTCTGCTGACTCTT 1619
 Db 1752 rAsnTrpGlnLysLeu 1757

RESULT 15

US-07-910-760-10

; Sequence 10, Application US/07910760

; Patent No. 5683864

; GENERAL INFORMATION:

; APPLICANT: Houghton, Michael

; APPLICANT: Choo, Qui-Lim

; APPLICANT: Kuo, George

; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)

TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: P.O. Box 8097 (Int. Prop. R-440)
CITY: Emeryville
STATE: CA
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/910,760
FILING DATE: 07-JUL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Blackburn Esq., Robert P.
REGISTRATION NUMBER: 30,447
REFERENCE/DOCKET NUMBER: 0101,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2702
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-910-760-10

Alignment Scores:

Pred. No.:	0.0139	Length:	3011
Score:	117.00	Matches:	91
Percent Similarity:	35.49%	Conservative:	46
Best Local Similarity:	23.58%	Mismatches:	138
Query Match:	3.06%	Indels:	111
DB:	1	Gaps:	23

US-09-743-237-3 (1-2134) x US-07-910-760-10 (1-3011)

QY	641	TCTGCTTTCCCC-TCGGGATCAACTCTCCAGGACCACCAAAATAACTTTGGCTGGG	697
DB	1424	ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro	1478
QY	698	TACTGTGACTGTTTGGCCAGTGGGACTTTTGCACAACTGCAATTTGTAATAATTGTTGC	757
DB	1444	Tyr-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysVal	1458
QY	758	AACAACTTG-----CATCATGATATTGAACGGTTTAAAGCCATT	796
DB	1459	ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro	1478
QY	797	AAGCATGCTTGTGGTAGAAATCCAGAAGCTTTCCAGCCAAAATAATGGGAAGGCCAAATG	856
DB	1479	GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro	1496
QY	857	GGC-----AATGTCAGCCCCAGCACAAAGGTCGCAACTGCGAGAGGTCAGGC	907
DB	1497	GlyIleTyrArgPheValAlaProGlyGlu-----ArgProSerGly	1510
QY	908	TGCCTGAAGAAAT-----TACTGCCAGTGTCTATGAGGCCCAATATATGTTCTTCTATT	961
DB	1511	MetPheAspSerSerValLeuCysGluCysTyrAspAla-----	1523
QY	962	TGCAAAATGCTTGGTGAATAATATGAAGAACCCAGAACGACAAAGACACTAATGAGC	1021
DB	1524	-----GlyCysAlaTrpTyrGluLeuThrPro---AlaGluThrThrValArg	1538
QY	1022	ATGCCAAACTACATCGAGCTGGAGTTTG-----	1051

Search completed: April 21, 2003, 12:05:13
Job time : 105.5 secs

DB	1539	LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTyr	1558
QY	1052	GAAGGC-----AGCCATTACTGCCACCACCAAGAAATTT	1084
DB	1559	GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln	1578
QY	1085	TCAGGA-----CTTCCAGAGATTTCAGTC-----ACGATAGGCGCTTCTCTCA	1126
DB	1579	SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGln	1598
QY	1127	TGCATCTCTCTGGGAGGTGG-----TGGAGGCCACATGCGCTGCTGCTGCTCAG	1177
DB	1598	nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg	1613
QY	1178	GCAGAGAGGCGCGAGAAAGACACTGCTCCAAAGTGCCTGGCAGACGAGATGATCTGGAG	1237
DB	1613	gLeuLysProThrLeuHisGlyProThrProLeuTyrArgLeu-----	1628
QY	1238	GAATTTTGAAGGTGCTTATTCACAGATTCTCCACACTGACT-----	1277
DB	1629	-----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrIle	1645
QY	1278	-----TTAAATCTAAGGATTGAAATAGGAGTAGAGTATAAAG	1315
DB	1645	eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValcylGln	1665
QY	1316	TGTGAATGCA-----TGTGATTTTGTCTTACTAGTAATCTCTAGTTTGAAGAGAT	1369
DB	1665	yValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysValIleVal-----	1683
QY	1370	GTTTAGGGGAACATGAGGCTGCTGCGACCAACACAGGCTCCCTGCATCCCTGGGC	1429
DB	1684	-----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArg	1698
QY	1430	CCAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTTC	1483
DB	1698	gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIle	1718
QY	1484	TCAGGAGGTGCTGCGCTGAGCATTTGTTGCTGCGCCAGAGAGAGAGCTTGGGTTCCC	1543
DB	1718	eGluGln-----GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu--	1735
QY	1544	ATAGTCTGGGAGAGTGTCTGCGAGGCGCGGAGGCGAGAGAGGCTTCCGAGAGCTC	1603
DB	1736	-----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh	1752
QY	1604	ACTCTGGTGCAGCTCTT	1619
DB	1752	rAsnTrpGlnLysLeu	1757

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:55:45 ; Search time 53 Seconds

(without alignments)
6089.058 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtcaaggcgaag.....aaaaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 577658

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet_n2p.model -DEV=xlp

-Q=/cgn2_1/USProoool/US09743237/runat_21042003_112211_1541/app_query.fasta_1.2311

-DB=PublishedApplications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100

-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US09743237 -CGCN 1.1.25 -runat_21042003_112211_1541

-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

```
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437	11.4	438	10	US-09-220-091-13
2	165	4.3	1057	9	US-10-189-971-6
3	158	4.1	1192	9	US-10-189-971-18
4	158	4.1	1207	9	US-10-189-971-20

5	158	4.1	1477	9	US-10-189-971-8
6	158	4.1	1535	9	US-10-189-971-14
7	158	4.1	1593	9	US-10-189-971-4
8	157.5	4.1	1251	9	US-10-189-971-16
9	157.5	4.1	1342	9	US-10-189-971-24
10	157.5	4.1	1512	9	US-10-189-971-10
11	157.5	4.1	1570	9	US-10-189-971-12
12	157.5	4.1	1628	9	US-10-189-971-2
13	137	3.6	4123	9	US-10-213-509-5
14	130.5	3.4	2743	9	US-10-037-182-36
15	130.5	3.4	3695	9	US-10-037-182-2
16	119	3.1	1036	9	US-09-887-527-60
17	118.5	3.1	1801	10	US-09-938-275-8
18	118	3.1	732	9	US-09-977-418-20
19	117	3.1	2894	9	US-10-044-995-23
20	117	3.1	2894	10	US-09-941-611-23
21	117	3.1	3011	10	US-09-916-359-2
22	117	3.1	3635	9	US-10-037-182-4
23	117	3.1	3635	10	US-09-845-583-2
24	116	3.0	1036	9	US-09-373-967-4
25	116	3.0	1036	9	US-10-028-072-142
26	116	3.0	1036	9	US-10-121-049-142
27	116	3.0	1036	9	US-10-123-904-142
28	116	3.0	1036	9	US-10-140-470-142
29	116	3.0	1036	9	US-10-175-746-142
30	116	3.0	1036	9	US-10-176-918-142
31	116	3.0	1036	9	US-10-176-921-142
32	116	3.0	1036	9	US-10-137-863-142
33	116	3.0	1036	9	US-10-140-474-142
34	116	3.0	1036	9	US-10-142-431-142
35	116	3.0	1036	9	US-10-143-114-142
36	116	3.0	1036	9	US-10-140-002-142
37	116	3.0	1036	9	US-10-142-419-142
38	116	3.0	1036	9	US-10-123-262-142
39	116	3.0	1036	9	US-10-142-423-142
40	116	3.0	1036	9	US-10-121-050-142
41	116	3.0	1036	9	US-10-141-755-142
42	116	3.0	1036	9	US-10-143-032-142
43	116	3.0	1036	9	US-10-123-108-142
44	116	3.0	1036	9	US-10-123-236-142
45	116	3.0	1036	9	US-10-123-261-142

ALIGNMENTS

RESULT 1
US-09-220-091-13
; Sequence 13, Application US/09220091
; Patent No. US20020064523A1
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; EARLIER FILING DATE: 1998-12-23
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; EARLIER APPLICATION NUMBER: 09/087,136
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Alignment Scores:
Pred. No.: 3.56e-29 Length: 438
Score: 437.00 Matches: 92
Percent Similarity: 56.95% Conservative: 35

Best Local Similarity: 41.26% Mismatches: 62
Query Match: 11.43% Indels: 34
DB: 10 Caps: 8

US-09-743-237-3 (1-2134) x US-09-220-091-13 (1-438)

```
Qy 698 TACTGTGACTGCTTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTAATAATTGTTGC 757
|||||
Db 191 TyrCysAspCysPheAlaAsnGlyGluPheCysArgAspCysAsnCysLysAspCysHis 210
|||||
Qy 758 AACCACTGCATCATATTAAGCGTTTAAAGCCATTAAAGCATCTCTGGTGAAGAAAT 817
|||||
Db 211 AsnAsnIleGluTyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 230
|||||
Qy 818 CCAGAAAGCTTTCCAGCAAAATTTGGG-----AAGGCCAAATTTGGCAATGTCAAGCCC 871
|||||
Db 231 ProAsnAlaPheLysProLysIleGlyLeuAlaArgGlyGlyIleThrAspIleGluArg 250
|||||
Qy 872 CAGCACAAAGGTCGCACTCAGGAGGTGAGCTGCTGGAAGAATTTACTGCGAGTGC 931
|||||
Db 251 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrCysGluCys 270
|||||
Qy 932 TATCAGGCCCAAAATATGCTTCTTATTTGCAAAATGCATTGTTGCCAAAAT----- 985
|||||
Db 271 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 290
|||||
Qy 986 -----TATGAAGAAAGC-----CCAGAACGAAGACACTA 1015
|||||
Db 291 ThrTyrArgMetThrArgTyrLysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeu 310
|||||
Qy 1016 ATGAGCATGCCAACTACATCAGACT-----GGAGGTTTG--- 1051
|||||
Db 311 MetSerLeuThrAsnAlaSerSerThrAlaThrProAspSerGlyProGlySerValVal 330
|||||
Qy 1052 -----GAAGCAGCCATTACCTGCCACCACGAATTTTCAGGACTTCCAAGATTC 1102
|||||
Db 331 ThrAspGluHisGlyAspSerPtyr---GluAspMetLeuLeuSerHisLysProLysVal 349
|||||
Qy 1103 AGTCACGATAGCGG-----CCCTCTCATCATCTCCCTGGGAGGTGGTGGAGGCCACA 1156
|||||
Db 350 GluMetAspProArgArgPheProTyrTyrMetThrAspGluValValGluAlaAla 369
|||||
Qy 1157 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1198
|||||
Db 370 ThrMetCysMetValAlaGlnAlaGluGluAlaLeuAsnTyrGluLysValGlnThrGlu 389
|||||
Qy 1199 CACTGCTCCTCAAGTCCCTGGCAGACAGATCATCTGAGGAATTTGGAAGTGTCTATCA 1258
|||||
Db 390 AspGluLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGlu 409
|||||
Qy 1259 CAGATTCTC 1267
|||||
Db 410 GlnMetIle 412
```

RESULT 2

US-10-189-971-6

Sequence 6, Application US/10189971

Publication No. US20030028907A1

GENERAL INFORMATION:

APPLICANT: Walke, D. Wade

APPLICANT: Scoville, John

APPLICANT: Turner, C. Alexander Jr.

TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynucleoti

TITLE OF INVENTION: Same

FILE REFERENCE: LEX-0360-USA

CURRENT APPLICATION NUMBER: US/10/189,971

CURRENT FILING DATE: 2002-07-03

PRIOR FILING DATE: 2001-07-03

PRIOR FILING DATE: 2001-07-03

PRIOR FILING DATE: 2001-08-29

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 4.0.

SEQ ID NO 6
LENGTH: 1057
TYPE: PRT
ORGANISM: homo sapiens
US-10-189-971-6

Alignment Scores:

Pred. No.: 1.33e-05 Length: 1057
Score: 165.00 Matches: 171
Percent Similarity: 27.67% Conservative: 62
Best Local Similarity: 20.31% Mismatches: 226
Query Match: 4.31% Indels: 383
DB: 9 Gaps: 48

US-09-743-237-3 (1-2134) x US-10-189-971-6 (1-1057)

```
Qy 16 CGAAGCTCGCGGGGGGCGACAGCGCGGGGAGCTCCTCGGGAGTACCCCGGGATCC 75
|||||
Db 24 ArgThrThrAlaAlaAlaValProLeuAlaGlyLys-----SerThrProAlaGlu 40
|||||
Qy 76 CAGAGCTCAGCGCGCTGGAGAGC----- 99
|||||
Db 41 ArgThrSerProThrProLeuThrProAlaValCysValAlaValSerGlyAsnValGln 60
|||||
Qy 100 -----TCGCGCTCTCGAGCGCCCGCGCGCCG-----CCTGCAACG 138
|||||
Db 61 CysLeuAlaArgArgCysValProLeuProCysProGluProValLeuLeuProGlyGlu 80
|||||
Qy 139 TGCACCTTCCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 198
|||||
Db 81 Cys-----CysProGlnCysProAlaAlaProAlaGlyCysProArgProGly 98
|||||
Qy 199 GCG-----CCTGGGCTCTGCGAAGGAGCGCTCC 225
|||||
Db 99 AlaAlaHisAlaArgHisGlnGluTyrPheSerProGlyAsp----- 113
|||||
Qy 226 CACCGCGGCTCCGATGATCCAGTTGAATCAAGTAAGCAGCGGTGCTACTACTACAAG 285
|||||
Db 114 ---ProCysArgArgCysLeuCysLeuAspGlySerValSer----- 126
|||||
Qy 286 TAATAATCCGGAAGAACCACTTTGCAGAACTCTTCTGCTCAGGAATCCTGTTCAAGTT 345
|||||
Db 127 -----CysGlnArgLeuProCysProProAlaProCysAlaHis---ProArgGlnGly 143
|||||
Qy 346 CCAATGCTCCAGGAACATAGAGGATGCTCTG-----CTGTTCTCTTAAGAA 393
|||||
Db 144 ProCysCysProSerCysAspGlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGlu 163
|||||
Qy 394 AGATTC---CAACCCAAATGTGATATGCCAATTTGAAAGGGGCGACACAAATGCTATAT 450
|||||
Db 164 ArgPheProSerProThr-AlaAlaCysHisLeu-----CysLeuCysTr 178
|||||
Qy 451 AGACAATTCAGAACAGAGAACTTAAAGCACTC-----CATTTGCTTCTCT 496
|||||
Db 178 pGluGlySerValSerCysGluProLysAlaCysAlaProAlaLeuCysProPheProAl 198
|||||
Qy 497 -----CAGTATCAAGATCAAAATATTATCT 522
|||||
Db 198 aArgGlyAspCysCysProAspCysAspGlyCysGluTyrLeuGlyGlySerTyrLeuSe 218
|||||
Qy 523 ACAGTCAGATGCTCCTAAACCA-----ATGACTGCTTTAGTAGGAGATT 567
|||||
Db 218 rAsnGlnGluPheProAspProArgGluProCysAsnLeuCysThrCysLeuGlyGlyPh 238
|||||
Qy 568 TTTGCCAGCATCAACAAATTAATCTCATTACACAACTTTCAGGAGCGCTTACCATC 627
|||||
Db 238 eValThrCys-----GlyArgArgProCys 246
|||||
Qy 628 GGTAGTCAACGGGTCTGCTTTCCCC----- 652
|||||
Db 246 sGluProProGlyCysSerHisProLeuIleProSerGlyHisCysCysProThrCysGl 266
|||||
Qy 653 -----TCGGGATCAACTCTTCCAGGACCA----- 676
```

Db 266 nGlyCysArgTyrHisGlyValThrAlaSerGlyGluThrLeuProAspProLeuAs 286
 QY 677 -CCAAAATAACCTTGGCTGGTACTGCTGCTGGTGGGAGGAGCTTTGGCAACAA 735
 Db 286 pProThrCysSerLeu-----CysThrCysGlnGluGlySerMetArgCysGlnLy 303
 QY 736 C-----TGCAATTGTAATAA 750
 Db 303 sLysProCysAlaProAlaLeuCysProHisProSerProGlyProCysPheCysProVa 323
 QY 751 TTGTTGCAACAACCTGCATCATGATATTGAACGGCTTTAAAGCCATTAAAGCCATGCTCTGG 810
 Db 323 lCys-----HisSerCysLeuSe 329
 QY 811 TAGAAATCCAGAACCTTTCAGCCAAAATTTGGGAAGGGCCCAATTGGGCAATGTCAGCC 870
 Db 329 rGlnGlyArgGluHisGlnAspGlyGluGluPheGluGlyProAlaGlySerCysGlu-- 348
 QY 871 CCAGCACAACAAAGGTGCAACTGCAGGAGTCAAGGCTGCTGAAGAATTACTGCGAGTG 930
 Db 349 -----trpCysArgCy 352
 QY 931 CTATGAGGCCCAATATATGT 952
 Db 352 sGlnAlaGlyGlnValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGl 372
 QY 953 -----TCCTCTATTTC-----AAATGCAATTGGTTCGCAAAATATAT---GAAGA 993
 Db 372 nValThrGluArgGlySerCysCysProArgGlnGlyProHisAspCysCysProGlnCys 392
 QY 994 AAGCCCAAGAAAGACACTAATGAGCATGCCAACTACATGCAGACTGCAGAGTTTGA 1053
 Db 392 uHisPro-----G 395
 QY 1054 AGGCAGCATTACCTGCCACACAGAAATTTTCAGGACTTCCAGATTTCAGTCACGATAG 1113
 Db 395 uGlySerArgTrpValPro-----AspSe 404
 QY 1114 GCGGCTTCTCTCATCATCTCTCTGGAGGTGGTGGGCGGCACATGCGCC-----TG 1164
 Db 404 rAlaCysSerSerCysValCysHisGlyVal---ValThrCysAlaArgIleGlnCy 423
 QY 1165 CCTG-----CTTGCTCAGGAGAGAGGCGGAGAAAGAACACTGCTCCAGAGTGCCTGGC 1218
 Db 423 sIleSerCysAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 441
 QY 1219 AGACCATGATCCTGGAGGAATTTGAAGTGCTTATACAGATTCTCCACATGAGTT 1278
 Db 442 -----SeraspCy 444
 QY 1279 TAAATCTAAGGATTGAAATGGAG-----TAGAGTATAAGTG 1317
 Db 444 sGluHisGluGlyArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCy 464
 QY 1318 TGAATGATGTTGATTGTTGCTTAGCTAGCTAGAAATCTCTAGTTTAGAAAGGATGTTAGGG 1377
 Db 464 sGluValCysIle----- 468
 QY 1378 GAACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCTGCATCCCTGGGCCCGAGGAG 1437
 Db 469 -----CysGluProGlnProGluGlyProProSerLeu----- 479
 QY 1438 TTTACTCAGACTCTCTGAAGATGT---GGCAACCCATGCCCTTTCTTGAGGA----- 1489
 Db 480 -----ArgCysHisArgGlnCysProSerLeuValGlyCysPr 493
 QY 1490 -----GGTGCATGCCCTGAGCATTTG----- 1510
 Db 493 oProSerGlnLeuLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLe 513
 QY 1511 -----TTGCTGGGCCCA----- 1522

Db 513 uSerAsnCysSerGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrTh 533
 QY 1523 -----GAGGAGAGAGCTGGGTTCCATA----- 1546
 Db 533 rCysGlnCysGlnAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCy 553
 QY 1547 -----GTCCTGGGAGAGTGTCT 1563
 Db 553 sProLeuSerGluArgHisThrProProGlySerCysCysProValCysArgGluCysVa 573
 QY 1564 GCAGGGCGGCGGAGGCGAGAGCCCTGGGAGAGCTCACTGTGTCGACTCTTCTC 1623
 Db 573 lVal-GluAlaGluGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProS 591
 QY 1624 TCAGAGATGT-----TGCTCTGGAGGC-----T 1647
 Db 591 erAsnAlaCysIleAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluC 611
 QY 1648 GCTCTGCATGAAAACCCCTAATGTTCTGTTCTGTTCTTCAAAATATTAGAAATAAGTT 1707
 Db 611 ySGLnAla-----LeuSerCysProHisGlyTrpAlaLysValProG 625
 QY 1708 CTCCGGATGGCTGTTGTGATACCATTAATAATCTTAGAGAACTACTGAACACTAAAG 1767
 Db 625 lAlaAsp-SerCysCysGluArg----- 632
 QY 1768 ATTTCTGTACCGTAGATATTTCCCA-----GAGACACCGGAACCTGTCAGTCTTT 1818
 Db 633 -----CysGlnAlaProThrGlnSerCysValHisGlnGlyArgGluVal----- 647
 QY 1819 CCTAAGGCCCGGAGGAGCAGGCAATGGGCGCTCGCAGGCCAGGCTTGCCAGCATG 1878
 Db 648 -----AlaSerGlyGluArg-----TrpThrVal-----AspThrCysThrSerCys 661
 QY 1879 TCTTGAGTTAGAGACTTAAATATCCAGTTTCTGCTGTTCTTCTGTTCTTCTGTTCTGGA 1938
 Db 662 SerCysMetalGlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 681

RESULT 3

US-10-189-971-18
 ; Sequence 18, Application US/10189971
 ; Publication NO. US20030028907A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walke, D. Wade
 ; APPLICANT: Scoville, John
 ; APPLICANT: Turner, C. Alexander Jr.
 ; TITLE OF INVENTION: No. US20030028907A1 Human Kiellin-like Proteins and Polynuc
 ; FILE REFERENCE: LEX-0360-USA
 ; CURRENT FILING DATE: 2002-07-03
 ; PRIOR FILING DATE: 2001-07-03
 ; PRIOR APPLICATION NUMBER: US 60/302,949
 ; PRIOR FILING DATE: 2001-08-29
 ; PRIOR APPLICATION NUMBER: US 60/315,634
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 18
 ; LENGTH: 1192
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-189-971-18

Alignment Scores:
 Pred. No.: 5,58e-05 Length: 1192
 Score: 158.00 Matches: 169
 Percent Similarity: 27.63% Conservative: 57
 Best Local Similarity: 20.66% Mismatches: 219
 Query Match: 4.13% Indels: 374
 DB: 9 Gaps: 46

US-09-743-237-3 (1-2134) x US-10-189-971-18 (1-1192)

```
QY 14 GCGAAGCTCCGGGGGCGACACGAGCGGGGGAGCTCTCGGGGAGTACCCCGGGAT 73
Db 198 GlyCysAlaPheGlyGlyLysGluTyrProSerGlyAlaAspPheProHisProSerAsp 217
QY 74 CCCAG-----AGCTCAGCGCGCTGGAGGAGCTCGCGCTCTCGCAGG 114
Db 218 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 237
QY 115 CCCCGCAGCCCGCG-----CCTGCAACGTGCACTTCTCTCTCGCTCGCTGC 159
Db 238 ProLeuProCysProGluProValLeuLeuProGlyGlyCys-----CysProGlnCys 255
QY 160 TACCCGCGCAGCCCGCGGGTGTGTCCTCCCTCGCGCG-----201
Db 256 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 275
QY 202 -----CCTGGGTCTCGGAAGGAGCTCCACCGCGCGCTCCGATGATC 246
Db 276 GluTyrPheSerProProGlyAsp-----ProCysArgArgCysLeu 289
QY 247 CCAGTTGAAATCAAGSTAAAGCAGGTGTACTACTACAAAGTAATAATCCGGAAGAAGCAAC 306
Db 290 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 301
QY 307 TTTCAGAAATCTTCTGCTCAGGAATCTGTTGCAAGTTCCTCATGTCCTCCAGGAAGTAGA 366
Db 302 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 320
QY 367 GGATGCTCTCTG-----CTGTTCTCTTAAGAAAGATTC---CAACCCCAATGGT 411
Db 321 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-AL 340
QY 412 GATATGCCAATTAAGAGGGGCGCACAAATGCTATGATAGACAAATTTAGAACAAAGAGA 471
Db 340 aAlaCysHisLeu-----CysLeuCysTyrGluGlySerValSerCysGly 355
QY 472 ACTAAAGCACTC-----CATTTGGTTCT-----496
Db 355 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 375
QY 497 -----CAGTATCAAGATCAAAATAATATTACTACATCGATGCTCCCTAAACC 543
Db 375 pCysAspGlyCysGluTyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspPr 395
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTCCAGCATCAACAAAT 588
Db 395 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys-----411
QY 589 AAATCTCATTACACAACTTCAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
Db 412 -----GlyArgProCysGluProProGlyCysSerH1 423
QY 649 CCCC-----652
Db 423 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 443
QY 653 -----TCGGGATCAACTTCCAGACACCACAAAATAACTTTGGCTGGGTACTG 702
Db 443 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr-----458
QY 703 TGACTGCTTTGCCAGTGGGACTTTTTCACAACTCAATTTGTAATAATTTGTCACAA 762
Db 459 -----CysSerLeuCysThrCysGln-----GlyAr 467
QY 763 CTTGCATCATGATPATTTGAACGGTTTAAAGCCATTAAAGCATGCTTGGTAGAAATCCAGA 822
Db 467 gGluHisGlnAspGlyGluPhe-----475
QY 823 AGCTTTCCAGCCAAAATTTGGGAGGGCAATTTGGCAATGTCAAGCCCGACCAACAA 882
Db 476 -----GluGlyProAlaGlySerCysGlu-----483
QY 883 AGGGTGCAACTGCAGGAGGTCCCTGAGAAATTACTGGAGTGTCTATGAGGCCCA 942
```

```
Db 484 -----TrpCysArgCysGlnAlaGlyG1 491
QY 943 AATTATGTGT-----952
Db 491 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 511
QY 953 -TCTTCTATTTC-----AAATGCAATTGGTTCAAAAATATAT---GAAGAAAGCCAGAACG 1005
Db 511 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro-----529
QY 1006 AAAGACACTAATAGCATGCCAAACTACATGCAGACTGGAGGTTTGGAGGCGAGCCATT 1065
Db 530 -----GluGlySerArgTr 534
QY 1066 CCTGCCACCAACCAAAATTTTCAGGACTTCCAAAGATTTCAGTCAGATAGGCGGCTTCCTC 1125
Db 534 pValProPro-----AspSerAlaCysSerSe 543
QY 1126 ATGCATCTCTCTGGAGGTGGTGGAGCCACATCGCG-----TGCTCTG-----CT 1170
Db 543 rCysValCysHisGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 562
QY 1171 TGCTCAGGAGAGAGCGCGAGAAAGAACTGCTCCAAAGTGCCTGGCAGAGCAGATGAT 1230
Db 562 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys-----576
QY 1231 CCTGGAGGAATTTGGAAGGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGG 1290
Db 577 -----SerAspCysGluHisGlu 583
QY 1291 ATTGAAAATGGAG-----TAGATATAAAGTGTGAATGCATGTT 1329
Db 583 yArgLysTyrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIl 603
QY 1330 GATTTGCTTAGTCTAGAAATCTAGTTTAGAAGGATGTTTAGGGGAACATGAGGCT 1389
Db 603 e-----603
QY 1390 GGCTCTGTCAGAACAAACAGGCTCCCTGCATCCCTGGCGCCAGGAGTTTACTCAGAGC 1449
Db 604 ---CysGluProGlnProGluGlyProProSerLeu-----614
QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCTTTTCTGAGGA-----1489
Db 615 ---ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 632
QY 1490 ---GGTGCATGCCCTGACCATGT-----1510
Db 632 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 652
QY 1511 ---TTGCTGCGCCA-----GA 1524
Db 652 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyrThrCysGlnCysGl 672
QY 1525 GGAGAGAGCTTGGTTTCCATA-----1546
Db 672 nasPLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGl 692
QY 1547 -----GTCCTGGGAGAGTGTCTCGAGGGCGCG 1575
Db 692 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAlaAg 712
QY 1576 AGGGCAGACAGCGCCTCGGAGAGCTCCTCTGTCGCTGCTCTCTCAGAGATGT- 1634
Db 712 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 730
QY 1635 -----TGCTCTGGAGGC-----TGCTCTGCATGAA 1659
Db 730 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla---748
QY 1660 AACCCTAATGTTTCTTGTGTTTCTTCAATATTTAGAAATAGTTCTCGGATGGGC 1719
```


Db 749 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 763
Qy 1720 TGTGTGATACCACTTAAATCTCTAGAGAACTACTGAACACCTTAAAGATTTTCTGTAGC 1779
|||||
Db 764 CysCysGluArg-----CysGln 769
Qy 1780 GTAGATATTTCCCA-----GAGACAGCGGAAGTGTCTTCTTCTTAAAGCCGCC 1830
|||
Db 770 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 784
Qy 1831 GGGAGACCGACGCAATGGGGCTCGCAGCCGAGGCTTGCACACGATGCTTGTAGTTAGA 1890
|||
Db 785 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMeta 800
Qy 1891 GCACTTAAATATCCAGCTTCTCTGTGTCTTCTTCTTCTTGAATTTGGA 1938
|||
Db 801 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 816
RESULT 4
US-10-189-971-20
; Sequence 20, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; TITLE OF INVENTION: Turner, C. Alexander Jr.
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189,971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ IDS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-20
Alignment Scores:
Pred. No.: 5,6e-05 Length: 1207
Score: 158.00 Matches: 169
Percent Similarity: 27.63% Conservative: 57
Best Local Similarity: 20.66% Mismatches: 219
Query Match: 4.13% Indels: 374
DB: Gaps: 46
US-09-743-237-3 (1-2134) x US-10-189-971-20 (1-1207)
Qy 14 GCGAAGCTCGGGGGCGACGACGAGCGGGAGCTCTCGGGAGTACCCCGGAT 73
|||
Db 213 GlyCysAlaPheGlyGlyLysGluTrpProSerGlyAlaAspPheProHisProSerAsp 232
Qy 74 CCCAG-----AGTCAGCGCGCTGGAGGAGCTCGCGTCTCTGAGG 114
|||
Db 233 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 252
Qy 115 CCGCGACCGCCCG-----CCTGCAAGCTGCACCTCTCTGCTCTCGCTGC 159
Db 253 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 270
Qy 160 TACCGCGCACCGCGCGCGGTGTTTTCCTCGGGCG-----201
Db 271 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 290
Qy 202 -----CCTGGGTCTCGGAAGGAGCTCCCGCGGGCTCCCGATGATC 246
Db 291 GlutyrPheSerProProGlyAsp-----ProCysArgArgCysLeu 304

Qy 247 CCAGTTGAAATCAAGGTAAGCAGGTGGTACTACTACAAGTAATAATCCGGAAGAAC 306
|||
Db 305 CysLeuaspGlySerValSer-----CysGlnArgLeuPro 316
Qy 307 TTTCAGAAATCTTCTGCTCAGGAATCTGTGCAAGTTTCCATGGTCCCGAAGACTAGA 366
|||
Db 317 CysProProAlaProCysAlaHis-----ProArgGlnGlyProCysCysProSerCysAsp 335
Qy 367 GGATGCTCTCCG-----CTGTTCTCTTAAGAAAGATTC-----CAACCCCAATGGT 411
|||
Db 336 GlyCysLeuTrpGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 355
Qy 412 GATATGCCAATTGAAAGGGGCACACAAATCTATGATAGACAAATTTCTAGAACAGAGA 471
|||
Db 355 AlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGln 370
Qy 472 ACTAAAGCACTC-----CATTTGGTTCT-----496
Db 370 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 390
Qy 497 -----CAGTATCAAGATCAAAATTAATTTCTACAGTCAGATGTCCTTAAC 543
Db 390 pCysAspGlyCysGluTrpLeuGlyLysSerTrpLeuSerAsnGlnGluPheProAspPr 410
Qy 544 A-----ATGACTGCTTTAGTAGGGAGATTTTTCAGCATCAACAAAAAT 588
Db 410 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys-----426
Qy 589 AAATCTCATTACACAACTTTGAGGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
Db 427 -----GlyArgArgProCysGluProProGlyCysSerHi 438
Qy 649 CCCC-----552
Db 438 sProLeuileProSerGlyHisCysCysProThrCysGlnGlyCysArgTrpHisGlyVa 458
Qy 653 -----TCGGGATCAACTCTTCCAGGACCAACCAAAATACTTTGGTGGTACTG 702
Db 458 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr-----473
Qy 703 TGACTGCTTTGCCAGTGGGAGCTTTTGCAACAACTGCAATTTGTAATAATTTGTCACAA 762
Db 474 -----CysSerLeuCysThrCysGln-----GlyAr 482
Qy 763 CTTCATCATCATATTGAACGTTTAAAGCCATTAAGCATCTCTTGTAGAAATCCAGA 822
Db 482 gGluHisGlnAspGlyGluGluPhe-----490
Qy 823 AGCTTTCAGCCAAAAATTTGGGAAGGGCAATTTGGGCAATTTGCAAGCCCCAGCACAA 882
Db 491 -----GluGlyProAlaGlySerCysGlu-----498
Qy 883 AGGTGCAACTGCAGAGGTGCTGCTGGAAGAAATTTACTCGAGTGTCTATGAGCCCA 942
Db 499 -----TrpCysArgCysGlnAlaGlyGln 506
Qy 943 AATTATGTGT-----952
Db 506 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 526
Qy 953 -TCTTCTATTTC-----AAATGCAATTTGCAAAATTAAT-----GAAGAAAGCCAGAACG 1005
Db 526 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro-----544
Qy 1006 AAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGGAAAGGAGCAGCATTA 1065
Db 545 -----GluGlySerArgTr 549
Qy 1066 CTGCCACCAACAAATTTTTCAGGACTTCCAGATTTCCAGATTCAGTACAGTACGGCGCTTCTC 1125
Db 549 pValProPro-----AspSerAlaCysSerSe 558
Qy 1126 ATGCATCTCTCTGGGAGGTGGTGGAGCCACATGCGCC-----TCCTG-----CT 1170


```
QY 497 -----CAGTATCAAGATCAAAATAATTATCTACAGTCAGATGTCCTCAACCC 543
Db 660 pCysAspGlyCysGluThrLeuGlyGluSerTyLeuSerAsnGlnGluPheProAspPr 680
QY 544 A-----ATGACTGCTTTAGTACGGAGATTTTCCAGCAGCATCAACAAATY 588
Db 680 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 696
QY 589 AAATCTCTATACAAACAACTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
Db 697 -----GlyArgArgProCysGluProProGlyCysSerHi 708
QY 649 CCCC----- 652
Db 708 sProLeuileProSerGlyHisCysCysProThrCysGlnGlyCysArgTyHisGlyVa 728
QY 653 -----TCGGATCACTCTCCAGGACCACCAAAATAACTTTGGCTGGGTACTG 702
Db 728 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThr----- 743
QY 703 TGAAGTCTTGGCAGTGGGACTTTTGCAACAACCTGCAATTCATAATTTGTCACAA 762
Db 744 -----CysSerLeuCysThrCysGln-----GlyAr 752
QY 763 CTTCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATCTCTTGTAGAAATCCAGA 822
Db 752 gGluHisGlnAspGlyGluGluPhe----- 760
QY 823 ACCTTCCAGCCAAATAATTGGAGGGCAATTTGGCAATGTCAAGCCCAACCAACAA 882
Db 761 -----GluGlyProAlaGlySerCysGlu----- 768
QY 883 AGGGTGCAACTGCAGGAGTCAGGCTGCTGAAGATTAAGTCTCGAGTGTATGAGGCCCA 942
Db 769 -----TrpCysArgCysGlnAlaGlyGly 776
QY 943 AATTATGTGT----- 952
Db 776 nValSerCysValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluAr 796
QY 953 -TCTTCTATTTCG---AAATGATTTGGTTGCAAAATAT---GAAGAAGCCCAAGG 1005
Db 796 gGlySerCysCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 814
QY 1006 AAAGACACTAATGAGCATGCCAACTACATGCAGACTGGAGCTTTGGAGGCGCAGCATTA 1065
Db 815 -----GluGlySerArgTr 819
QY 1066 CCGTCCACCAACAAATTTTCAGGACTTCCAAAGATTTCAGTCAGTACGATAGCGGCTTCCTC 1125
Db 819 pValProPro-----AspSerAlaCysSerSe 828
QY 1126 ATGCATCTCTGGAGGTGGTGGAGCCACATGCGGC-----TGCTG-----CT 1170
Db 828 rCysValCysHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCy 847
QY 1171 TGCTCAGGAGAGAGCGCGAGAAAGAACACACTGCTCCAAAGTCCCTGGCAGACCATGAT 1230
Db 847 sAlaGlnProArgGlnGlyProHisAspCysCysProGlnCys----- 861
QY 1231 CCGGAGGAATTTGGAAGGTGCTTATCAGAGATTCTCCACATGATTTAAATTAAGG 1290
Db 862 -----SerAspCysGluHisGluGly 868
QY 1291 ATTGAATGGAG-----TAGAGTATAAGTGTGAATGATGTT 1329
Db 868 yArgLysTyThrGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysI 888
QY 1330 GATTTTGTCTAGTCTAGAAATCTCTAGTTTAGAAAGGATGTTTAGGGGAACATGAGGCT 1389
Db 888 e----- 888
QY 1390 GGCTCTGCAGCAACCAACCGAGGCTCCCTGCTATCCCTGGGCCCGAGGATTTACTCAGAGC 1449
```

```
Db 889 ----CysGluProGlnProGluGlyProProSerLeu----- 899
QY 1450 TCTCTGAAGATGT---GGCAACCCATGCCCCCTTTTCTAGGA----- 1489
Db 900 -----ArgCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLe 917
QY 1490 ----GGTGCATGGCTGAGCATTTG----- 1510
Db 917 uLeuProProGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSe 937
QY 1511 -----TTGCTGTGCCCA-----GA 1524
Db 937 rGluGlyLeuLeuGlySerGluLeuAlaProProAspProCysTyThrCysGlnCysGl 957
QY 1525 GGAGAGAGCTTGGGTCCCAT----- 1546
Db 957 nAspLeuThrTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGl 977
QY 1547 -----GTCTGGGAGAGTGTCTCAGGGCGCGG 1575
Db 977 uArgHisThrProProGlySerCysCysProValCysArgGluCysValVal-GluAla 997
QY 1576 AGGCGAGAGCAGCGCTCGGAGAGCTCACTCTGGTCTGACTCTCTCTCAGAGAATGT- 1634
Db 997 luGlyArgArg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysI 1015
QY 1635 -----TGCTCTGGAGGC-----TGCTCTGCATGAA 1659
Db 1015 leAlaCysThrCysHisArgGlyHisValGluCysHisLeuGluCysGlnAla----- 1033
QY 1660 AACCTAATGGTTTCTGTTGTTTTCATAATTAATTAAGTTTCCGGATGGGC 1719
Db 1034 -----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-Ser 1048
QY 1720 TGTGTGTATACCACTTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTTCTGTAGC 1779
Db 1049 CysCysGluArg-----CysGln 1054
QY 1780 GTAGATATTTCCTCCA-----GAGACACCGCAACTCTCAGTCTTCTAAGGCCCCC 1830
Db 1055 AlaProThrGlnSerCysValHisGlnGlyArgGluVal-----AlaSer 1069
QY 1831 GGGACAGCAGGCAATGGGCGCTCGCAGGCCAGGCTTGACACGAGCATGCTTTGAGTTAGA 1890
Db 1070 GlyGluArg-----TrpThrVal-----AspThrCysThrSerCysSerCysMetAla 1085
QY 1891 GGACTTAAATTTATCCAGTTTCTCTGTGTTTCTTACTTGAATTTGTGA 1938
Db 1086 GlyThrValArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1101
```

RESULT 6

```
US-10-189-971-14
; Sequence 14, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucle
; FILE REFERENCE: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302, 949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315, 634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1535
; TYPE: PRT
```


Qy	367	GGATGCTCTCTG-----CTGTTCTCTTAAGAAAGATTG---CAACCCCAATGGT	411
Db	345	GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-AL	364
Qy	412	GATATGCCAATTAAAGGGGCACAAATGCTATGTATAGACAATTCCTAGACAACAGAGA	471
Db	364	aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysG	379
Qy	472	ACTAAAGCACTC-----CATTTGGTTCTCT-----	496
Db	379	uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs	399
Qy	497	-----CAGTATCAGATCAAAATAATTATCTACAGTCAGATGTCCTTAACCC	543
Db	399	pCysAspGlyCysGluTyrLeuGlyCysTyrLeuSerAsnGlnGluPheProAspPr	419
Qy	544	A-----ATGACTGCTTTAGTAGGGAGATTTTTTTCGCAGCATCAACAAAATT	588
Db	419	oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys-----	435
Qy	589	AAATCTATTACACAACACTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTT	648
Db	436	-----GlyArgArgProCysGluProProGlyCysSerHi	447
Qy	649	CCCC-----	652
Db	447	sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa	467
Qy	653	-----TCGGGATCAACTCTTCCAGGACCA-----CCAAAAATAACTTTGGCTGG	696
Db	467	lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----	485
Qy	697	GTACTGTGACTGCTTTGCCAGTGGGACCTTTTGGCAACAC-----	736
Db	486	----CysThrCysGlnGluGlySerMetArgCysLysLysProCysAlaProAlaLe	504
Qy	737	-----TGCAATTGTAATAATTGTTGTCACAACTTGCATCA	771
Db	504	uCysProHisProSerProGlyProCysPheCysProValCys-----	518
Qy	772	TGATATTGAACGGTTAAAGCCATTAAAGCATGCTTGGTAGAAATCCAGAAGCTTTCCA	831
Db	519	-----HisSerCysLeuSerGlnGlyArgGluHisGlnAs	530
Qy	832	GCCAAAATTGGGAAGGGCCATTGGCAATGCTCAAGCCCCCAGCACACAAAGGTCGAA	891
Db	530	pGlyGluGluPheGluGlyProAlaGlySerCysGlu-----	542
Qy	892	CTGCAGGAGGTACGGCTGCCTGAAGAATTTACTGCGAGTGTATGAGGCCCAATATTGTG	951
Db	543	-----TrpCysArgCysGlnAlaGlyGlnValSerCy	553
Qy	952	T-----TCTTCTAT	960
Db	553	sValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCy	573
Qy	961	TTGC---AAATGCTATTGGTAAAAAATTAT---GAAAGAACCCAGACACAAAGACACT	1014
Db	573	sCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro-----	588
Qy	1015	AATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGAAGGCGAGCCATTACCTGCCACC	1074
Db	589	-----GluGlySerArgTrpValProPr	596
Qy	1075	AACGAATAATTTCAGGACTTCCAAAGATTACATCAGATAGCGGCGCTTCCTCATGCATCTC	1134
Db	596	o-----AspSerAlaCysSerSerCysValCy	605
Qy	1135	CTGGAGGTGGTGAGGCCACATGCGCC-----TGCCTG-----CTTGCTCAGGG	1179
Db	605	sHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr	624

[illegible]

```
: APPLICANT: Walke, D. Wade
: APPLICANT: Scoville, John
: APPLICANT: Turner, C. Alexander Jr.
: TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotides
: TITLE OF INVENTION: Same
: FILE REFERENCE: LEX-0360-USA
: CURRENT APPLICATION NUMBER: US/10/189,971
: CURRENT FILING DATE: 2002-07-03
: PRIOR APPLICATION NUMBER: US 60/302,949
: PRIOR FILING DATE: 2001-07-03
: PRIOR APPLICATION NUMBER: US 60/315,634
: PRIOR FILING DATE: 2001-08-29
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 24
: LENGTH: 1342
: TYPE: PRT
: ORGANISM: homo sapiens
: US-10-189-971-24

Alignment Scores:
Pred. No.: 6,42e-05 Length: 1342
Score: 157.50 Matches: 172
Percent Similarity: 27.78% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 231
Query Match: 4.12% Indels: 373
DB: 9 Gaps: 47

US-09-743-237-3 (1-2134) x US-10-189-971-24 (1-1342)
Qy 14 GCGGAAGCTCGGGGCGGACGACGCGGGGGAGTCTCTCGGGAGTACCCGGGAT 73
Db 313 GlyCysAlaPheGlyGlySerProSerGlyAlaAspPheProHisProSerAsp 332
Qy 74 CCCAG-----AGTCACGCGCTGGAGGAGTCCGCGTCTCTCGAGG 114
Db 333 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgCysVal 352
Qy 115 CCCGCGACGCCCG-----CCTCAAGCTGCACCTCTCTCGTCTCGTGC 159
Db 353 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 370
Qy 160 TACCCGCGACCCGCGCGGGTGTTCCTCCCTCGGGCG----- 201
Db 371 ProAlaAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 390
Qy 202 -----CCTGGTCTCGCAAGGAGCCTCCACCGCGCGTCCGATGATC 246
Db 391 GluTyrPheSerProGlyAsp-----ProCysArgArgCysLeu 404
Qy 247 CCAGTTGAATCAAGGTAAGCAGGTGGTACTACTACAAGTAATAATCCGGAAGCAAC 306
Db 405 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 416
Qy 307 TTGCAGAACTCTTGTCTGCTGAGTAATCCTGTGCAAGTTCCTGTCCTCCAGAACTAGA 366
Db 417 CysProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 435
Qy 367 GGATGCTCTCCG-----CTGTTCTTAAAGAAATTC---CAACCCCAATGGT 411
Db 436 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-AL 455
Qy 412 GATATGCCAAATTGAAGGGGCGACACAATGCTATGATACAAATCTAGAACAAAGAGA 471
Db 455 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGln 470
Qy 472 ACTAAACCACTC-----CATTTGGTTCCT----- 496
Db 470 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 490
Qy 497 -----CAGTATCAAGTCAAAATAATTACTACAGTCAGATGTCCTTAACC 543
Db 490 pCysAspGlyCysGluTyrLeuGlyGlySerTyrLeuSerAsnGlnGluPheProAspPr 510
```

```
Qy 544 A-----ATGACTGCTTTAGTAGGAGATTTTTCAGCATCAACAAAAT 588
Db 510 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 526
Qy 589 AAATCTCATTACACAACAACCTTACAGGAGCCTTACCATCGGTAGTCAACGGGTGCTTT 648
Db 527 -----GlyArgProCysGluProGlyCysSerHis 538
Qy 649 CCCC----- 652
Db 538 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 558
Qy 653 -----TCGGGATCAACTCTTCAGGACCA-----CCAAAAATAACTTTGGCTGG 696
Db 558 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----- 576
Qy 697 GTACTGTGACTGCTTTGCCAGTGGGACCTTTTGCACAAC----- 736
Db 577 ---CysThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLe 595
Qy 737 -----TGCATTTGTAATAATTTTGCACAACTTGCATCA 771
Db 595 uCysProHisProSerProGlyProCysPheCysProValCys----- 609
Qy 772 TGATATTGAACGGTTTAAAGCCATTAAAGGCATGCTTGGTAGAAATCCAGAAAGCTTTCCA 831
Db 610 -----HisSerCysLeuSerGlnGlyArgGluHisGlnAs 621
Qy 832 GCCAAAAATGGGAAGGCCAAATTTGGGCAATGTCAAGCCCAAGCACAACAAAGGTCGAA 891
Db 621 pGlyGluGluPheGluGlyProAlaGlySerCysGlu----- 633
Qy 892 CTGAGGAGGTGAGCTGCTCGCTGAAGAATTTACTGCGAGTGTCTATGAGCCCAAAATATGTG 951
Db 634 -----TrpCysArgCysGlnAlaGlyGlnValSerCys 644
Qy 952 T-----TCTTCTAT 960
Db 644 sValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCys 664
Qy 961 TTGC---AAATGCAATGTTGCAAAAATAT---GAGAAAGCCCAAGCAACGAAAGACT 1014
Db 664 sCysProArgCysArgGlyCysLeuAlaHisGlyGluHisPro----- 679
Qy 1015 AATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGAAGGCGAGCCATTACCTGCCACC 1074
Db 680 -----GluGlySerArgTrpValPro 687
Qy 1075 AACGAAATTTTCAGGACTTCCAAGATTACGTACAGTAGGCGCTTCTCATCATCTC 1134
Db 687 o-----AspSerAlaCysSerSerCysValCys 696
Qy 1135 CTGGGAGTGTGTGAGGCCACATCGGC-----TGCCTG-----CTTGTCTCAGG 1179
Db 696 sHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 715
Qy 1180 AGAAGAGCCGAGAAAGAACACTGCTCAAGTGCCTGCAGAGCAGATGATCCTGGAGGA 1239
Db 715 oArgGlnGlyProHisAspCysCysProGlnCys----- 726
Qy 1240 ATTTGGGAAGTGTCTATACAGATTCTCCACACTGAGTTTAAATCTAAGGATTGAAAAT 1299
Db 727 -----SerAspCysGluHisGluGlyArgLysTy 736
Qy 1300 GGAG-----TAGAGTATAAGTGAATGCAATGCTGTTGATTTGCT 1338
Db 736 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 753
Qy 1339 TTAGTCTAGAAATCTCTAGTTTAGAAAGGATGTTTAGGGGAACATGAGGCTGGCTCTCA 1398
Db 754 -----CysGln 755
```


Qy	1534	TTGGGTTCCCAT-	154
		:	
Db	995	rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArGHisTh	1015
Qy	1547	-----GTCTGGGAGAGTCTCTGGTCAGCTCTCTCTCTCAGAGAATGT	1584
		:	
Db	1015	rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA	1035
Qy	1585	CAGGCCCTGGGGAGAGCTCACCTGTGTCTGCTCAGCTCTCTCTCTCAGAGAATGT	1634
		:	
Db	1035	rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCysT	1053
Qy	1635	--TGCTCTGGAGGC-----TGCTTCGCATGAAACCCCTAAT	1668
Db	1053	hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla-	1068
Qy	1669	GGTTTCTGTTCTGTTTTCAAATATTATTAGAATAAGTTCTCCCGATGGGCTGTTGTGAT	1728
		: : :	
Db	1069	----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-SerCysCysGlu	1086
Qy	1729	ACCATTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTCCTGTAGCGTAGATATT	1788
Db	1087	Arg-----CysGlnAlaProThr	1092
Qy	1789	TCCCCA-----GAGACACGCAACTGTGAGTCTTCTCTAAGGCCCGGGAGACGC	1839
		: :	
Db	1093	GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg	1107
Qy	1840	AGCAATGGGCTCCAGCCAGGCTGCACCACGATGTTCTGAGTTAGAGGACTTAA	1899
		: :	
Db	1108	-----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal	1123
Qy	1900	ATTATCCAGTTCTCTGTGTTTCTACTTGAATTGTGA	1938
Db	1124	ArgCysGlnSerGlnArgCysSerProLeuSerCysGly	1136
RESULT 11			
US-10-189-971-12			
; Sequence 12, Application US/10189971			
; Publication No. US20030028907A1			
; GENERAL INFORMATION:			
; APPLICANT: Walke, D. Wade			
; APPLICANT: Scoville, John			
; APPLICANT: Turner, C. Alexander Jr.			
; TITLE OF INVENTION: No. US20030028907A1el Human Kielin-like Proteins and			
; TITLE OF INVENTION: Same			
; FILE REFERENCE: LEX-0360-USA			
; CURRENT APPLICATION NUMBER: US/10/189,971			
; CURRENT FILING DATE: 2002-07-03			
; PRIOR APPLICATION NUMBER: US 60/302,949			
; PRIOR FILING DATE: 2001-07-03			
; PRIOR APPLICATION NUMBER: US 60/315,634			
; PRIOR FILING DATE: 2001-08-29			
; NUMBER OF SEQ ID NOS: 25			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 12			
; LENGTH: 1570			
; TYPE: PR1			
; ORGANISM: homo sapiens			
US-10-189-971-12			
Alignment Scores:			
Pred. No.:		6,77e-05	Length: 1570
Score:		157.50	Matches: 172
Percent Similarity:		27.78%	Conservative: 60
Best Local Similarity:		20.60%	Mismatches: 231
Query Match:		4.12%	Indels: 373
DB:		9	Gaps: 47
US-09-743-237-3 (1-2134) x US-10-189-971-12 (1-1570)			
Qy	14	GGCGAAGCTCGCGGGGGGACAGCAGCGCGGGGAGTCTCTCGGGGAGTACCCCGGAT	73
		:	

Db 541 GlyCysAlaPheGlyGlyLysGluTyProSerGlyAlaAspPheProHisProSerAsp 560
QY 74 CCAG-----AGCTCAGCGCGCTGGAGGACGTCGCGCTCCCTGCAGG 114
Db 561 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgArgCysVal 580
QY 115 CCCCGAGCGCGCG-----CTGCAAGGTGCACCTCCCTGCTCCGCTGC 159
Db 581 ProLeuProCysProGluProValLeuLeuProGlyGluCys-----CysProGlnCys 598
QY 160 TACCGCGCGCAGCGCGCGGGTGTTCGCCCTGGGCG----- 201
Db 599 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 618
QY 202 -----CCTGGGTCTCGGAAGGAGCTCCACCGCGGCTCCGCGATGATC 246
Db 619 GluTyPheSerProProGlyAsp-----ProCysArgArgCysLeu 632
QY 247 CCAGTTGAATCAAGGTAAAGAGGTGGTACTACTACAAGTAATAATCCGGAAGAAGCAAC 306
Db 633 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 644
QY 307 TTTCAGAAATCTTCTGCTCAGAAATCTGTTGCAAGTTCCTCCATGGTCCCGAAGACTAGA 366
Db 645 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 663
QY 367 CGATGCTCTCTG-----CTGTCTCTTAAGAAAGATTC---CAACCCAAATGCT 411
Db 664 GlyCysLeuTyGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThrAl 683
QY 412 GATATGCCAAATTAAGAGGGGCGACAAATGCTATGATAGACAAATTTCTAGAACAGAGA 471
Db 683 aAlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGln 698
QY 472 ACTAAAGCAGCTC-----CATTTGGTTCCT----- 496
Db 698 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 718
QY 497 -----CAGTATCAAGATCAAAATAATATCTACAGTCAGATGTCCTCAAAACC 543
Db 718 pCysAspGlyCysGluTyLeuGlyGlySerTyLeuSerAsnGlnGluPheProAspPr 738
QY 544 A-----ATGACTGCTTTAGTAGGAGATTTTGGCAGCATCAACAAATTT 588
Db 738 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 754
QY 589 AAATCTCATTACACACAACTTGAGGGAGCCTTACCATCGGTAGTCACGGGTCTGCTTT 648
Db 755 -----GlyArgArgProCysGluProProGlyCysSerHi 766
QY 649 CCCC----- 652
Db 766 sProLeuIleProSerGlyHisCysCysProThrCysGlnGlyCysArgTyHisGlyVa 786
QY 653 -----TCGGGATCAACTTTCAGGACCA-----CCAAATAAATTTGGCTGG 696
Db 786 lPhrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----- 804
QY 697 GTACTGTGACTCTTTCGACGTGGGAGCTTTTGGCAACAC----- 736
Db 805 -----CysThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLe 823
QY 737 -----TGCAATTTGTAATAATTTGTTGCAACAACTTCATCA 771
Db 823 uCysProHisProSerProGlyProCysPheCysProValCys----- 837
QY 772 TGATATTGAACGGTTTAAAGCAATTAAGCATGCTTGGTAGAATAATCCAGAACTTTCCA 831
Db 838 -----HisSerCysLeuSerGlnGlyArgGluHisGlnAs 849
QY 832 GCCAAAATTTGGGAAGGCCAATTCGGCAATCTCAAGCCCCAGCACAAAGGTGCA 891
Db 849 pGlyGluGluPheGluGlyProAlaGlySerCysGlu----- 861

QY 892 CTGCAGGAGGTGAGGCTGCCTGAAGAATTACTTCCGAGTGTCTATGAGGCCCAAAATATGTG 951
Db 862 -----TrpCysArgCysGlnAlaGlyGlnValSerCy 872
QY 952 T-----TCTTCTAT 960
Db 872 sValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCy 892
QY 961 TTGC---AAATGCATTGTTGCAAAATAT---GAAGAAAGCCCAAGACGAAAGACACT 1014
Db 892 sCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 907
QY 1015 AATGAGATGCCAAACTACATCAGACTGGAGGTTTGAAGGAGGCACATTAACCTGCACCC 1074
Db 908 -----GluGlySerArgTrpValProPr 915
QY 1075 AACGAAATTTTCAGGACTTCCAAGATTACGTACCATAGGCGGCTTCTCTCATCATCTC 1134
Db 915 O-----AspSerAlaCysSerSerCysValCy 924
QY 1135 CTGGGAGGTGGTGGAGGCCACATGCGCC-----TCCTG-----CTTGTCTCAGG 1179
Db 924 sHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 943
QY 1180 AGAAGAGCGCGAGAAAGAACACTGCTCCAAGTGCCTGGCAGAGCAGATGATCCTGGAGGA 1239
Db 943 oArgGlnGlyProHisAspCysCysProGlnCys----- 954
QY 1240 ATTTGGAAGGTCTTATCACAGATTCTCCACACTGAGTTAAATCTAAGGATTAAGAAAT 1299
Db 955 -----SerAspCysGluHisGluGlyArgLysTy 964
QY 1300 GGAG-----TAGAGTATAAGTGTGAATGCATGCTTGTGATTTGTC 1338
Db 964 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle----- 981
QY 1339 TTAGTCTAGAAATCTCTAGTTTAGAAGGATGTTTAGGGGAACATGAGCTGCTCTGCA 1398
Db 982 -----CysGln 983
QY 1399 GCAACAACAGGCTCCCTGCTCATCCCTGGGCCAGGAGTCTTACTCAGAGCTCTCTGAAG 1458
Db 983 uProGlnProGluGlyProProSerLeu-----Ar 993
QY 1459 ATGT---GGCAACCATGCCCTTTTCTGAGGA-----GGTGC 1494
Db 993 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 1013
QY 1495 ATGGCTGTGACATTTGT----- 1510
Db 1013 oGlyProGlnHisCysCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 1033
QY 1511 -----TTGTCTGGCCCA-----GAGGAGAGAGC 1533
Db 1033 uLeuGlySerGluLeuAlaProProAspProCysTyThrCysGlnCysGlnAspLeuTh 1053
QY 1534 TTGGGTTCACATA----- 1546
Db 1053 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 1073
QY 1547 -----GTCTCTGGAGAGTGTCTCGAGGCGCGGAGGAGGAGAG 1584
Db 1073 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA 1093
QY 1585 CAGGCCCTCGGAGAGCTCCTGCTGCTGCTTCTCTCTCAGAGAATGT----- 1634
Db 1093 rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCyst 1111
QY 1635 --TGCTCTGGAGGC-----TGCTCTGATGAAACCCCTAAT 1668
Db 1111 hrCysHisArgGlyHisValGluCysHisLeuGluGluCysGlnAla----- 1126

```
QY 1669 GGTTCCTGTTGTTTCAAAATATTAGAAATAAGATTCTCCGGATGGGTGTGTGAT 1728
Db 1127 ----LeuSerCysProHisGlyTrpAlaLysValProGlnAlaAsp-SerCysCysGlu 1144
QY 1729 ACCACTTAAATCTCTAGACACTACTGACACACCTAAAGATTCTCTAGGCTAGATATT 1788
Db 1145 Arg-----CysGlnAlaProThr 1150
QY 1789 TCCCCA-----GAGACACGGAACTGTCAGTCTTCTCCTAAGGCCCGCGGAGACGC 1839
Db 1151 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 1165
QY 1840 AGGCAATGGGCGCTCCGAGGCCAGCTTGACACAGCATGTCTTGAGTTAGAGCACTTAAA 1899
Db 1166 ----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 1181
QY 1900 ATTATCAGTTCTCTGTTCTGTTCTACTTGAATTGTGGA 1938
Db 1182 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1194

RESULT 12
US-10-189-971-2
; Sequence 2, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kiellin-like Proteins and Polynucleotid
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302, 949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315, 634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1628
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-2

Alignment Scores:
Pred. No.: 6.86e-05 Length: 1628
Score: 157.50 Matches: 172
Percent Similarity: 27.78% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 231
Query Match: 4.12% Indels: 373
DB: 9 Gaps: 47

US-09-743-237-3 (1-2134) x US-10-189-971-2 (1-1628)
QY 14 GCCGAAGCTCGCGGGGCGACACGACGCGGGGAGCTCCTCGGGGAGTACCCCGGAT 73
Db 599 GlyCysAlaPheGlyGlyLysGlyTrpSerGlyAlaAspPheProHisProSerAsp 618
QY 74 CCCAG-----AGCTACGCGCTGGAGGAGCTCGCGCTCCTCGCAGG 114
Db 619 ProCysArgLeuCysArgCysLeuSerGlyAsnValGlnCysLeuAlaArgArgCysVal 638
QY 115 CCCCGAGCCCGCG-----CCTGCACGCTGCCTTCTGCTCCTCGCTGC 159
Db 639 ProLeuProCysProGluProValLeuLeuLeuProGlyGluCys-----CysProGlnCys 656
QY 160 TACCCGCGCACCGACCGCGCGGTGTTTGCCTCCCTCGGGCG----- 201
Db 657 ProAlaAlaProAlaProAlaGlyCysProArgProGlyAlaAlaHisAlaArgHisGln 676
QY 202 -----CCTGGGTCTCGCAAGAGGAGCCCTCCACCCGCGCGCTCCCGCATGC 246
Db 973 -----GluGlySerArgTrpValProPr 973
```

```
Db 677 GluTyrPheSerProGlyAsp-----ProCysArgArgCysLeu 690
QY 247 CCAGTTGAATCAAGGTAAACAGGTGTGTACTACTACAAGTAATAATCCGGAAGAACAC 306
Db 691 CysLeuAspGlySerValSer-----CysGlnArgLeuPro 702
QY 307 TTTGCGAATCTTCTTCTGCTCAGGAATCCTGTTGCAAGTTCCTCCATGTTCCCGAGCACTAGA 366
Db 703 CysProProAlaProCysAlaHis---ProArgGlnGlyProCysCysProSerCysAsp 721
QY 367 GGATGCTCTCTG-----CTGTTCTTTAAGAAAGATTTC-----CAACCCCAATGGT 411
Db 722 GlyCysLeuTyrGlnGlyLysGluPheAlaSerGlyGluArgPheProSerProThr-Al 741
QY 412 GATATGCCAATTCGAAGGGGCGACACAAATGCTATGATAGACAATCTAGACAACAGA 471
Db 741 AlaCysHisLeu-----CysLeuCysTrpGluGlySerValSerCysGly 756
QY 472 ACTAAAGCACTC-----CATTTGGTTCCT----- 496
Db 756 uProLysAlaCysAlaProAlaLeuCysProPheProAlaArgGlyAspCysCysProAs 776
QY 497 -----CAGTATCAAGATCAAATAATATCTACAGTCAGATGTCCTTAACC 543
Db 776 pCysAspGlyCysGluTyrLeuGlyGluSerTyrLeuSerAsnGlnGluPheProAspPr 796
QY 544 A-----ATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAACAAAATT 588
Db 796 oArgGluProCysAsnLeuCysThrCysLeuGlyGlyPheValThrCys----- 812
QY 589 AAATCTCATTACACAACAACCTTACGAGGCTTACCATCGGTAGTCAACGGTCTGCTTT 648
Db 813 -----GlyArgArgProCysGluProProGlyCysSerH1 824
QY 649 CCCC----- 652
Db 824 sProLeuLeuProSerGlyHisCysCysProThrCysGlnGlyCysArgTyrHisGlyVa 844
QY 653 -----TCGGATCAACTCTTCCAGGACCA-----CCAAAAATAACTTTGGCTGG 696
Db 844 lThrThrAlaSerGlyGluThrLeuProAspProLeuAspProThrCysSerLeu----- 862
QY 697 GTACTGTGACTGCTTGGCCAGTGGGACTTTTGCACAAC----- 736
Db 863 ---CysThrCysGlnGluGlySerMetArgCysGlnLysLysProCysAlaProAlaLe 881
QY 737 -----TCGAATGTAATAATTGTCACAACTTGCATCA 771
Db 881 uCysProHisProSerProGlyProCysPheCysProValCys----- 895
QY 772 TGATATTGAACGGTTTAAAGCCATTAAAGGCATGCTTGTGGTAGAAATCCAGAAGCTTTCCA 831
Db 896 -----HisSerCysLeuSerGlnGlyArgGluHisGlnAs 907
QY 832 GCCAAAAATTGGGAAGGGCCAATTGGCAATGTCAGCCCAACCAACAAAGGGTCAA 891
Db 907 pGlyGluGluPheGluGlyProAlaGlySerCysGlu----- 919
QY 892 CTCAGGAGCTCAGGCTGCTGCTGAAGAATTACTGCGAGTGTCTATGAGCCCAAAATATGCTG 951
Db 920 -----TrpCysArgCysGlnAlaGlyGlnValSerCys 930
QY 952 T-----TCTTCTAT 960
Db 930 sValArgLeuGlnCysProProLeuProCysLysLeuGlnValThrGluArgGlySerCys 950
QY 961 TTGC-----AAATGCTGTTGTCAAAAATTTAT---GAAGAAGCCCAACGAACAACT 1014
Db 950 sCysProArgCysArgGlyCysLeuAlaHisGlyGluGluHisPro----- 965
QY 1015 AATGAGCATGCCAAACTACATGACAGCTGGAGGTTTGGGAAGGAGCCATTTACCTGCCACC 1074
Db 966 -----GluGlySerArgTrpValProPr 973
```

```
QY 1075 AACGAAATTTTCAGGACTTCCAGATTTCAGTCACGATAGCGGCTTCCTCATGCATCTC 1134
Db 973 O-----AspSerAlaCysSerSerCysValCy 982

QY 1135 CTGGAGGTGGTGGAGCCACATGGCC-----TCCTG-----CTTCTCAGAG 1179
Db 982 sHisGluGlyVal---ValThrCysAlaArgIleGlnCysIleSerSerCysAlaGlnPr 1001

QY 1180 AGAAGAGCCGAGAGAAGACACTGCTCCAAAGTGCCTGGCAGACAGATCCTGGAGGA 1239
Db 1001 oArgGlnGlyProHisAspCysProGlnCys-----1012

QY 1240 ATTGGAAGGTCTTATCACAGATTCTCACACTGAGTTTAATCTAAGCGATTGAAAT 1299
Db 1013 -----SerAspCysGluHisGluGlyArgIleTy 1022

QY 1300 GGAG-----TAGAGTATAAGTGTGAATGCATGCTTCAATTTGTC 1338
Db 1022 rGluProGlyGluSerPheGlnProGlyAlaAspProCysGluValCysIle-----1039

QY 1339 TTAGTCTAGAAATCTCTAGTTTAGAAGGATGTTTAGGGGAACATCAGGCTGGCTCTGCA 1398
Db 1040 -----CysG1 1041

QY 1399 GCAACAACAGGCTCCCTGTCATCCCTGGCCAGGAGCTTTACTCAGAGCTCTCTGAAG 1458
Db 1041 uProGlnProGluGlyProProSerLeu-----Ar 1051

QY 1459 ATGT---GGCAACCATGCCCCCTTTTCTGAGGA-----GGTGC 1494
Db 1051 gCysHisArgArgGlnCysProSerLeuValGlyCysProProSerGlnLeuLeuProPr 1071

QY 1495 ATGGCTGAGCATTTGT-----1510
Db 1071 oGlyProGlnHisCysCysProThrCysAlaGluAlaLeuSerAsnCysSerGluGlyLe 1091

QY 1511 -----TGTCTGGGCCA-----GAGGAGAGAGC 1533
Db 1091 uLeuGlySerGluLeuAlaProAspProCysThrCysGlnCysGlnAspLeuth 1111

QY 1534 TTGGTTCCCAT-----1546
Db 1111 rTrpLeuCysIleHisGlnAlaCysProGluLeuSerCysProLeuSerGluArgHisTh 1131

QY 1547 -----GTCTGGGAGAGTGTCTGCAGGGCGGAGGAGCAGAG 1584
Db 1131 rProProGlySerCysCysProValCysArgGluCysValVal-GluAlaGluGlyArgA 1151

QY 1585 CAGGCCCTGGGAGAGCTCACTCTGTCGACTCTTCTCTCAGAGAATGT-----1634
Db 1151 rg---ValAlaAspGlyGluSerTrpArgAsp---ProSerAsnAlaCysIleAlaCysT 1169

QY 1635 --TGCTCTGGAGC-----TGCTCTGCATGAAACCCCTAAT 1668
Db 1169 hrCysHisArgGlyHisValGluCysHisLeuGluCysGlnAla-----1184

QY 1669 GCTTCTCTGTTGTTTTTCAAAATATTAGAAATAGTTCTCCGGATGGCTGTTGTGAT 1728
Db 1185 -----LeuSerCysProHisGlyTrpAlaIalysValProGlnAlaAsp-SerCysCysGlu 1202

QY 1729 ACCACTTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTCCTGTAGCGCTAGATAT 1788
Db 1203 Arg-----CysGlnAlaProThr 1208

QY 1789 TCCCCA-----GAGACACGCAACTGTCAGTCTTTCCTAAGCCCCCGGAGAGCC 1839
Db 1209 GlnSerCysValHisGlnGlyArgGluVal-----AlaSerGlyGluArg 1223

QY 1840 AGGCAATGGGGCTCGCAGGCCAGGCTGCACACAGCATGCTTGAGTTAGAGGACTTAA 1899
Db 1224 -----TrpThrVal-----AspThrCysThrSerCysSerCysMetAlaGlyThrVal 1239
```

```
QY 1900 ATTATCCAGTTTCTGTGTTTCTTACTTGAATTGTGGA 1938
Db 1240 ArgCysGlnSerGlnArgCysSerProLeuSerCysGly 1252

RESULT 13
US-10-213-509-5
; Sequence 5, Application US/10213509
; Publication No. US20030054485A1
; GENERAL INFORMATION:
; APPLICANT: Weiss, Joseph
; APPLICANT: Scott, Matthew
; TITLE OF INVENTION: JELLY BELLY GENES AND THEIR USES
; FILE REFERENCE: STAN-232
; CURRENT APPLICATION NUMBER: US/10/213,509
; CURRENT FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: 60/311,720
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 4123
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-213-509-5

Alignment Scores:
Pred. No.: 0.00552 Length: 4123
Score: 137.00 Matches: 153
Percent Similarity: 27.03% Conservative: 50
Best Local Similarity: 20.37% Mismatches: 208
Query Match: 3.58% Indels: 340
DB: Gaps: 44

US-09-743-237-3 (1-2134) x US-10-213-509-5 (1-4123)
QY 7 GGGTCAAGCGAAGCTCGCGGGG-----GCGACAGCAGCGCGGGGAGC 51
Db 2820 GlyAspArgGlnGluLeuGlnGlyCysHisThrValCysGlyThrGlyIleAlaGlySer 2839
QY 52 TCCTCGGGAGTACCCCGGATCCAGACTCA-----84
Db 2840 LeuGlyAlaGlyValProProSerSerSerGlnPheCysThrLeuArgThrHisGlyMet 2859
QY 85 -----GCGCGC 90
Db 2860 GlyProThrAspHisSerThrTrpGlyIleGluValPheGlyThrThrProThrSer 2879
QY 91 TGGAGAGC---TCGGCTCTGC-----AGGCCCGC 120
Db 2880 TrpSerSerCysSerGlnSerCysLeuAlaProGlyGlyGlyProGlyTrpArgSerArg 2899
QY 121 AGCCGC-----126
Db 2900 SerArgLeuCysProSerProGlyAspSerSerCysProGlyAspAlaThrGlnGlu 2919
QY 127 -----CCGCTCGAACGTCGACT-----TCCTGTCTCT---153
Db 2920 ProCysSerProProIleGluCysThrGlyPheCysAlaProGlyCysThrCysProPro 2939
QY 154 -----CGCTGCTACCCCGCACCGCCCGCGGTGT-----186
Db 2940 GlyLeuPheLeuHisAsnAlaSerCysLeuProArgSerGlnCys---ProCysGlnLeu 2958
QY 187 -----TTTGCCCTGGGCGCTGGG-----TCCTGCCAAGGAGCTCC 225
Db 2959 HisGlyGlnLeuTyAlaSerGlyAlaMetAlaArgLeuAspSerCysAsnAsnCysThr 2978
QY 226 CACCCGGGCGCTCCGATG-----ATCCAGTTCAAATCAAGGTAAGC 267
Db 2979 CysValSerGlyIleCysMetAlaCysThrSerGluArgCysProValAlaCysGlyTrpSer 2998
QY 268 AGGTGGTACTACTCAAGTAATAATCCGGAAGAACCACTTTGCAGAAATCTTCTTGCTCA 327
```

Db 2999 ProTrpThrLeuTrp-----SerLeuCysSer 3007
Qy 328 GGAATCCTGT----- 337
Db 3008 Cys-SerCysAsnValGlyIleArgArgPheArgAlaGlyThrAlaProProAlaAl 3027
Qy 338 -----TGCAGTTTCCCATGGTCCAGGAACACTAGAGATGGCTCTCTGCTG 381
Db 3027 aPheGlyAlaGluCysGlnGlyPro-----ThrMetGluAlaGluPheCy 3043
Qy 382 TTCCTCTTAAGAAAGATTCCCAACCAATGGTGCATATGCCAATTGAAAGGGGGCACACAAT 441
Db 3043 sSerLeuArgProCysProGlyProValProGlyMetCysProArgAspLysGlnTrpLe 3063
Qy 442 GCTATGTATAGACAAT---TCTAGAACAGAGAACTAAAGCACTCCATTGGTTCTCTCA 498
Db 3063 uAspCysAlaGlnGlyProAlaSerCysAlaGluLeuSerAla-----ProAr 3079
Qy 499 GTATCAAGATCAAAATAATTATCTACAGTCAGATGTCCTAAACCAATGACTGCTTTAGT 558
Db 3079 gGlyThrAsnGlnThrCysHisProGlyCysHisCysProSerGlyMetLeuLeuLeuVa 3099
Qy 559 AGGG---AGATTTTTCAGCATCAACAAATAAATCTATTACACAACTTGAGGG 615
Db 3099 lSerProArgGlyHisProGlyProLeuGlyAlaSerVal-----GlnProProVa 3116
Qy 616 AGCCTTACCATCGGTAGTCAAGGGTCTGCTTTCCTCCCTCGGATCA----- 661
Db 3116 lAlaLeuProGlyAlaIleGlyThrGlySerValProGlyAlaGlyGlyTrpGlyProTr 3136
Qy 662 -----AC 663
Db 3136 pGlyProTrpSerHisCysSerArgSerCysGlyGlyLeuArgSerArgThrArgAl 3156
Qy 664 TCITCCAGGACCACAAATAACTTTGGTGGTACTGTGAC---TGCITTCACAGTG 720
Db 3156 aCysAspGlnProProGlnGlyLeuGlyAspTrpCysGluGlyProArgAlaGlnGl 3176
Qy 721 GGACTTTTTC-----AACAACTGCAAT----- 742
Db 3176 yGluValCysGlnAlaLeuProCysProValThrAsnCysThrAlaIleGluGlyAlaGl 3196
Qy 743 -----TGTATAATTTGTGCAACAACATGTCATCATGATATGA 780
Db 3196 uTyrSerProCysGlyProProCysProArgSerCysAspLeuValHis----- 3213
Qy 781 ACGTTTAAAGCCATTAAAGCATCTCTGGTAGA---AATCCAGAAGCTTTCAGCCAAA 837
Db 3214 -----CysValTrpArgCysGlnProGlyCysTrpCysPro-- 3225
Qy 838 AATTGGGAAGGCCAATTTGGCAAT-----GTCAAGCCCCAGCACAA 879
Db 3226 -----ProGlyGlnValLeuSerSerAsnGlyAlaIleCysValGlnProGlyHis-- 3242
Qy 880 CAAAGGTGCAACTGC----- 895
Db 3243 -----CysSerCysLeuAspLeuLeuThrGlyGlnArgHisProGlyAlaArgLe 3260
Qy 896 ---AGAGGTTCAGCTGCCTGAGAGATTAATCTGCGAGTGTATGAGCCCAAAATATGTG 951
Db 3260 uAlaArgProAspGlyCys-----AsnHisCysThrCysLeuGluGlyArgLeuAsnCy 3278
Qy 952 TTCCTCTATTGCAATGCAATTTGTTGTCACAAAATTTATGAAGAACGCCAGAACGAAAGAC 1011
Db 3278 sThrAspLeu---ProCysProAspCys----- 3286
Qy 1012 ACTAATGAGCATGCCAAACTACATGCAGACTGGAGTTTGAAGGAGCCCATTTACCTGCC 1071
Db 3287 -----GlyGlyGlnSerLeuHis----- 3293
Qy 1072 ACCAAGCAAAATTTTCAGGACTTCACAGATTCAGTCAGATAGGGCGCTCTCCCTCA---TG 1128
Db 3294 -ProCysGly---GlnProCysProArgSerCysGlnAspLeuSerProGlySerValCy 3312

Qy 1129 CATCTCTCTGGAGGTG---GTGGAGGCCACATCGCCTGCTGCTGCTCAGGAGAAGA 1185
Db 3312 sGlnProGlySerValGlyCysGlnProThrCysGlyCysProLeuGlyGln----- 3329
Qy 1186 GCGCGAGAAAGAACACTGCTCCCAAGTCCCTGGCAGACAGATGATCTCTGGAGGAATTTGG 1245
Db 3330 -----LeuSerGlnAspGl 3334
Qy 1246 AAGTGCTTTATCAGAGATTCTCCACACTGAGTTTAAATCTTAAGGGATTGAAATGGAGTA 1305
Db 3334 yLeuCysValPro----- 3338
Qy 1306 GAGTATAAAAGTGAATGCATGTTGATTTTCTTCTTAGTCTAGAAATCTCTAGTTTAGAAA 1365
Db 3339 -ProAlaHisCysArgCys----- 3344
Qy 1366 GGATGTTTGGGGAACATGAGGTGGCTCTGCAGCACACCAACAGGCTCCCTGCA---TC 1422
Db 3345 -----GlnTyrGlnProGlyAlaMetAlaProSe 3354
Qy 1423 CCTGGGCCCGAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAAC-----CCATGCC 1476
Db 3354 rPheValProSerThrCysValAlaGlyIleLeuGlnCysGlnGluValProaspCysPr 3374
Qy 1477 CCTTTTCTGAGGAGTGCATGGCTCAGCATTTGTTCTCTGCCCCAGAGAGAGAGCTTG 1536
Db 3374 o-----AspProGlyValTrpSerSerTrp-----GlyProTrpGluAspCysSe 3389
Qy 1537 GGTTCCTCCATAGTCTCTGGAGAGTGTCTGAG----- 1567
Db 3389 rValSerCysGlyGlyGlyGluGlnLeuArgSerArgCysAlaArgProProCysPr 3409
Qy 1568 -GCGCGGGAGGAGCAGAGCAGAGCCCTCGCGAGAGCTCACTCTGCTGACTCTTCTCTCTCA 1626
Db 3409 oGlyProAlaArgGlnSerArgThrCys----- 1621
Qy 1627 GAGAATTTGCTCTGAGGAGCTGCTCTGCA 1655
Db 3421 n-ValCysArgGluAlaGlyCysProAla 3430

RESULT 14

US-10-037-182-36
; Sequence 36, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thybøll, Jili
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2743
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-037-182-36

Alignment Scores:
Pred. No.: 0.0174 Length: 2743
Score: 130.50 Matches: 111
Percent Similarity: 31.14% Conservative: 55
Best Local Similarity: 20.83% Mismatches: 198
Query Match: 3.41% Indels: 170
Gaps: 28
DB:

```
US-09-743-237-3 (1-2134) x US-10-037-182-36 (1-2743)
Qy 18 AAGCTCGGGGGGACAGCAGCGGGGAG-----CTCCTCGGGAGTAC 65
Db 83 LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleArgGlyGlnTyr 102
Qy 66 CCCGGATC-----CCAGAGCTCAGCGCGCTGGAG 95
Db 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleasp 122
Qy 96 GACGTC-----CGGCTCTGCGAGCCCGCCGACCCCGCCGCTGC 134
Db 123 GlyThrGluArgTrpTrpGlnSerProLeuSerArgGlyLeuGluTyrAsnGluVal 142
Qy 135 AACGTG-----CACTTCTCTGCTCGCTGCTACCCGCGCAC 170
Db 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 162
Qy 171 CGCAGCCCGGGGTGTTTGGCCCTGGCGCGCTGGTCTGCGAGGAGGCTCCACCC 230
Db 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetaspPhe 178
Qy 231 GGGGCTCGCGATGATCCAGTTGAATCAAGTAGCAGGTGTTACTACTACAAGTAATA 290
Db 179 GlyArgThrTyrGlnPro-----TrpGlnPhePhe 188
Qy 291 ATCCGGAGAAGCAACTTTGCGAATCTTCTGCTCAGGAATCTGTTGCAAGTCCCAT 350
Db 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199
Qy 351 GGTCCCGAGAACTA-----GAGGATGCTCTGCTGCTTCTCTTAAGAAA 394
Db 200 GlyProGlnThrLeuGluArgIleThrArgAspAspAlaAlaIleCysThrThrGluTyr 219
Qy 395 GATCCCAACCCATGGTGATATGCAATGAAGGGGCGACACAATGCTATGATATAGAC 454
Db 220 SerArg-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal 235
Qy 455 AATCTAGAACAAAGAACTAAAGCACTCCATTTGGTTCTCAGTATCAAGATCAAAAT 514
Db 236 AsnGlyArgProGlyAlaMet----- 242
Qy 515 AATATCTACAGTCAGATGCCCTAAACCAATGACTGCTTTAGTAGGG-----AGA 565
Db 243 AsnPheSerTyrSerProLeuLeuArgGluPheThrLysAlaThrAsnValArgLeuArg 262
Qy 566 TTTTTCAGCAGTCAACAATAA---AATCTATTACACAACACTTGGAGGACCTTA 622
Db 263 PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp----- 280
Qy 623 CCATCGGTAGTCAACGGGTCTGCTTCCCTCGGGATCACTCTTCCAGGACCAACAAA 682
Db 281 ProThrValThrArgArgTyrTyr-----SerIleLysasp 293
Qy 693 ATAATTTGGTGGTACTGTGACTGCTTTCCAGTGGGGACTTTTGC----- 730
Db 294 IleSerIleGlyArgCysValCysHisGlyHisAlaAspAlaCysAspAlaLysasp 313
Qy 731 -----ACAACCTGC-----AATGCT 745
Db 314 ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyGlyThrCys 333
Qy 746 AATAATCTTGCACAACTTCATCATGATATTGAACGGTTTAAAGCCATTAAAGCA--- 802
Db 334 AspArgCysCysProGlyPheAsnGln-----GlnProTrpLysProAlaThrAlaAsn 351
Qy 803 -----TGTCTTGTGTAGA-----AAT 817
Db 352 SerAlaAsnGluCysGlnSerCysAsnCysTyrGlyHisAlaThrAspCysTyrTyrAsp 371
Qy 818 CCAGAA-----GCTTCCAGCCAAAATTTGGAGGGGCCAATTTGGGCAAT 862
Db 372 ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyrGlnGlyGlyGly 391
```

```
Qy 863 GTC-----AAGCCCCAGCACAAAGGGTCAACTGCAGGAGTGCAGGCTGCCTG 913
Db 392 ValCysIleAspCysGlnHisThrAlaGlyValAsnCysGluArg-----CysLeu 409
Qy 914 AAGAATTACTCGAGTGTATGAGCCCAATTATGTCTTCTTATTGTTGAAATGCATT 973
Db 410 ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgArgCys 429
Qy 974 GGTTCAAAAT-----TATGAAGAAACCCAGACGAAGAACACTA 1015
Db 430 AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys 447
Qy 1016 ATGAGCATGCCAAACTAC-----ATGCAGACTGGAGGTTTGAAGGC 1057
Db 448 TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly 467
Qy 1058 AGCCATTACCTGCCAACCAAGAAATTTTCAGGACTTCCAAGATTTCAGTACGATAGCGG 1117
Db 468 PheProSerCysTyrProThr-----ProSerSerSerAsnAspThrArg 482
Qy 1118 CCTTCTCTCATCATCTCTGGAGGTGTGGAGGCCACATGGCCTGCTGCTGCTCAG 1177
Db 483 GluGlnValLeuProAlaGlyGlnIleValAsnCysAspCysSerAlaAlaGlyThrGln 502
Qy 1178 GGAGAAGAGGCGGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGACAGATGATCCTGGAG 1237
Db 503 GlyAsnAlaCysArgLysasp-----Pro 510
Qy 1238 GAATTTGGAAGTGTCTTATCA-----CAGATTCTCCACACTGAGTTTAAATCT 1285
Db 511 ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla 530
Qy 1286 AAGGATTGAATGGAGTACAGTAGATAAAGTGTGAATGC 1324
Db 531 ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys 543
```

RESULT 15

```
US-10-037-182-2
; Sequence 2, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thyboll, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3695
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-037-182-2
```

```
Alignment Scores:
Pred. No.: 0.0193 Length: 3695
Score: 130.50 Matches: 111
Percent Similarity: 31.14% Conservative: 55
Best Local Similarity: 20.83% Mismatches: 198
Query Match: 3.41% Indels: 170
DB: 9 Gaps: 28
```

US-09-743-237-3 (1-2134) x US-10-037-182-2 (1-3695)

```
Qy 18 AAGCTCGGGGGGACAGCAGCGGGGAG-----CTCCTCGGGAGTAC 65
Db 18 AAGCTCGGGGGGACAGCAGCGGGGAG-----CTCCTCGGGAGTAC 65
```

Db 83 LysLeuValGlyGlyProValAlaGlyGlyAspProAsnGlnThrIleArgGlyGlnTyr 102
QY 66 CCGGGATC-----CCAGAGCTCAGCGCGCTGGAG 95
Db 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleAsp 122
QY 96 GACGTC-----GCCTCTCGAGCGCGCGCGCGCGCTGC 134
Db 123 GlyThrGluArgTrpTrpGlnSerProProLeuSerArgGlyLeuGluTyrAsnGluVal 142
QY 135 AACGTG-----CACTTCCTGCTCGCTGCTACCGCGGCAC 170
Db 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 162
QY 171 CGCAGCGCGGGGTGTTTCCCTGGCGGCTCGGCTCGGAGGAGCGCTCCACCC 230
Db 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe 178
QY 231 GGGGTCGGCATGATCCAGCTTGAATCAAGTAAGCAGGTGTACTACTCAAGTAATA 290
Db 179 GlyArgThrTyrGlnPro-----TrpGlnPhePhe 188
QY 291 ATCGGAAGAAGCAACTTTCAGAACTCTTCTGCTCAGGAATCCTGTGCAAGTTCCTCAT 350
Db 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199
QY 351 GGTCCCGAGGAACATA-----GAGGATGCTCTCTGCTGCTCTTAAGAAA 394
Db 200 GlyProGlnThrLeuGluArgIleThrArgAspAlaAlaIleCysThrThrGluTyr 219
QY 395 GATTCCAACCCATGTGATATGCCAATTCGAAGGGGGCACAAATGCTATGTATAGAC 454
Db 220 SerArg-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal 235
QY 455 AATTCTAGAACAGAGAACTAAAGCACTCCATTGTTGTTCTCAGTATCAAGATCAAAAT 514
Db 236 AsnGlyArgProGlyAlaMet-----242
QY 515 AATTATCTACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGG-----AGA 565
Db 243 AsnPheSerTyrSerProLeuLeuArgGluPheThrLysAlaThrAsnValArgLeuArg 262
QY 566 TTTTGGCAGCATCAACAATAAFTA---AATCTCATTTACACAACAACTTGAGGGAGCCTTA 622
Db 263 PheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaLeuArgAsp-----280
QY 623 CCATCGGTAGTCAACGGGTCTGCTTTCCCTCGGGATCACTCTCCAGGACCAACCAAAA 682
Db 281 ProThrValThrArgArgTyrTyr-----SerIleLysAsp 293
QY 683 ATAATTTGGCTGGTACTGTGACTGCTTTGCCAGTGGGACTTTTGC-----730
Db 294 IleSerIleGlyGlyArgCysValCysHisGlyHisAlaAspAlaCysAlaLysAsp 313
QY 731 -----AACAACTGC-----AATTGT 745
Db 314 ProThrAspProPheArgLeuGlnCysThrCysGlnHisAsnThrCysGlyGlyThrCys 333
QY 746 AATAATTTGTCAACAACCTTGATATGTAACGGTTTAAAGCCATTAAGGCA---802
Db 334 AspArgCysCysProGlyPheAsnGln-----GlnProTrpLysProAlaThrAlaAsn 351
QY 803 -----TGTCTTGTGTAGA-----AAT 817
Db 352 SerAlaAsnGluCysGlnSerCysAsnCysTyrGlyHisAlaThrAspCysTyrTyrAsp 371
QY 818 CCAGAA-----GCTTTCCAGCCAAATAATTGGGAGGCCCAATTGGCCAAT 862
Db 372 ProGluValAspArgArgAlaSerGlnSerLeuAspGlyThrTyrGlnGlyGlyGly 391
QY 863 GTC-----AAGCCCCAGCACAAAGGGTGCACCTGCAGGAGGTCAAGCTGCCTG 913
Db 392 ValCysIleAspCysGlnHisHisThrAlaGlyValAsnCysGluArg-----CysLeu 409

QY 914 AAGAATTACTGCGAGTGCTATGAGGCCAAATATGTTGTTCTTCTATTGTCAAATGCATT 973
Db 410 ProGlyPheTyrArgSerProAsnHisProLeuAspSerProHisValCysArgArgCys 429
QY 974 GGTGCAAAAAT-----TATGAAGAAAGCCAGAACGAAAGACACTA 1015
Db 430 AsnCysGluSerAspPheThrAspGlyThrCysGluAspLeuThrGlyArg-----Cys 447
QY 1016 ATGAGCATGCCAAACTAC-----ATGCAGACTGGAGGTTTGGGAAGGC 1057
Db 448 TyrCysArgProAsnPheSerGlyGluArgCysAspValCysAlaGluGlyPheThrGly 467
QY 1058 AGCCATTACCTGCCACCAACGAAATTTTCAGACTTCCAGATTTCAGTCACGATAGCGG 1117
Db 468 PheProSerCysTyrProThr-----ProSerSerSerAsnAspThrArg 482
QY 1118 CCTTCCTCATGCTCCTCGGAGGTGGTGAGGCCACATGCGCTGCTGCTGCTCAG 1177
Db 483 GluGlnValLeuProAlaGlyGlnIleValAsnCysAspCysSerAlaAlaGlyThrGln 502
QY 1178 GGAGAAGAGCGCCGAGAAAGAACACTGCTCCAAGTGCTGCAGAGCAGATGATCCTGGAG 1237
Db 503 GlyAsnAlaCysArgLysAsp-----Pro 510
QY 1238 GAATTTGGAAGTGCTTATCA-----CAGATTCTCCACACTGAGTTTAAATCT 1285
Db 511 ArgValGlyArgCysLeuCysLysProAsnPheGlnGlyThrHisCysGluLeuCysAla 530
QY 1286 AAGGGATTGAAATGGAGTAGATATAAGTGTGAATGC 1324
Db 531 ProGlyPheTyrGlyProGlyCysGlnProCysGlnCys 543

Search completed: April 21, 2003, 12:06:48
Job time : 116 secs

GenCore version 5.1.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:51:19 ; Search time 65.5 seconds

(without alignments)
6264.148 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtgaagcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09743237/runat_21042003_112210_1486/app_query.fasta_1.2311
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237@cgn_1_1_111_@runat_21042003_112210_1486 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	437	11.4	429	T23152	hypothetical prote
2	376	9.8	571	B84585	hypothetical prote
3	358	9.4	603	T08955	hypothetical prote
4	248	6.5	658	F71410	hypothetical prote
5	123	3.2	1700	S08167	Balbani ring 3 pr
6	121	3.2	1737	T00209	MEGF8 protein - hu
7	119.5	3.1	1001	T00532	probable cadmium-t
8	119.5	3.1	1172	F84572	probable cadmium-t
9	119.5	3.1	3010	1 GNVVCJ	genome polyprotein
10	119	3.1	1895	T15881	hypothetical prote
11	118.5	3.1	1801	1 MMRIS	laminin beta-2 cha
12	118.5	3.1	5376	T42215	zonadhesin - mouse
13	118	3.1	2406	A54148	odz protein - frui
14	118	3.1	2515	S47008	tenascin-like prot

15	117.5	3.1	3507	2	T34513	hypothetical prote
16	117	3.1	497	2	T27827	hypothetical prote
17	117	3.1	3011	1	GNWVC3	genome polyprotein
18	117	3.1	3635	1	T10053	laminin alpha 5 ch
19	116	3.0	3010	1	GNWVC3	genome polyprotein
20	116	3.0	3010	1	A45573	genome polyprotein
21	115.5	3.0	769	1	IJHULM	leukocyte adhesion
22	115.5	3.0	1766	2	A42125	trophozoite cystei
23	115.5	3.0	2195	2	T34264	hypothetical prote
24	114.5	3.0	778	2	I38487	tastin - human
25	114.5	3.0	3759	2	A35085	trithorax protein
26	113.5	3.0	328	2	G89152	protein C24B5.5 [l
27	113	3.0	1820	2	A55494	latent transformin
28	112.5	2.9	574	2	B88465	protein B0244.8 [l
29	112.5	2.9	3133	2	S52093	hemocytin - silkw
30	112	2.9	1599	2	T16210	hypothetical prote
31	112	2.9	3010	1	GNWVTW	genome polyprotein
32	111	2.9	798	2	A40526	integrin beta-7 ch
33	110	2.9	1819	2	A71928	cag island protein
34	109.5	2.9	573	2	A33533	cell surface glyco
35	109.5	2.9	3010	1	S18030	genome polyprotein
36	109	2.9	2180	2	T29764	hypothetical prote
37	109	2.9	2499	1	A30788	mannose 6-phosphat
38	108.5	2.8	1548	2	S34583	serine proteinase
39	108.5	2.8	2225	2	T26063	hypothetical prote
40	108	2.8	1607	2	T02837	long chain fatty a
41	108	2.8	1840	2	T03250	Gr1 protein - mous
42	108	2.9	5376	2	T42215	zonadhesin - mouse
43	107	2.8	559	2	A57474	extracellular matr
44	106.5	2.8	1786	1	MMHUB1	laminin beta-1 cha
45	106.5	2.8	1797	2	A55677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T23152
hypothetical protein JC8.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Lightning, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19699
A:Accession: T23152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-429 <WIL>
A:Cross-references: EMBL:282274; PIDN:CAB05228.1; GSPDB:GN000022; CESP:JC8.6b
A:Experimental source: Clone JC8
C:Genetics:
A:Gene: CESP:JC8.6b
A:Map position: 4
A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Alignment Scores:
Pred. No.: 4,59e-28 Length: 429
Score: 437.00 Matches: 92
Percent Similarity: 56.95% Conservative: 35
Best Local Similarity: 41.26% Mismatches: 62
Query Match: 11.43% Indels: 34
DB: 2 Gaps: 8

US-09-743-237-3 (1-2134) x T23152 (1-429)

QY 698 TACTGTGACTGCTTTGACAGTGGGACTTTTGCACAACTGCAATTGTAATAATTGTTGC 757
|||||
Db 182 TyrcysaspCysphealaasnGlyGluPheCysArgaspCysasnCysLysAspCysHis 201
|||||
QY 758 AACAACTTGCATCATATTGACCGTTTAAAGCCCATGCTCTTGGTAGAAAT 817
|||||
Db 202 AsnAsnIleGlutyrAspSerGlnArgSerLysAlaIleArgGlnSerLeuGluArgAsn 221
|||||

```
Qy 818 CCAGAAGCTTCCAGCCCAAAATTTGG-----AAGGCCCAATTGGCAATGTCAAGCC 871
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 222 ProAsnAlaPheLysProLysIleGlyIleAlaArgGlyGlyIleThrAspIleGluArg 241
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 872 CAGCACAAAGGTCGAACCTCAGAGGTCAGCGTCCTCGAAGATTTACTGCGAGTGC 931
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 242 LeuHisGlnLysGlyCysHisCysLysSerGlyCysLeuLysAsnTyrCysGluCys 261
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 932 TATGAGGCCCAAAATATGTTCTTCTATTGCAAAATGCATTGGTGCACAAAT----- 985
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 262 TyrGluAlaLysValProCysThrAspArgCysLysCysLysGlyCysGlnAsnThrGlu 281
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 986 -----TATGAAGAAAGC-----CCAGAACGAAAGACACTA 1015
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 282 ThrTyrArgMetThrArgTyrLysAsnSerGlyGlyAlaValSerAsnThrAsnAlaLeu 301
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1016 ATGAGCATGCCAAACTACATCCAGACT-----GGAGGTTTG--- 1051
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 302 MetSerLeuThrAsnAlaSerSerThrAlaThrProAspSerGlyProGlySerValVal 321
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1052 -----GAAGCAGCCATTACCTGCCACCAAGAAATTTTCAGGACTTCCAGATTC 1102
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 322 ThrAspGluHisGlyAspTyr---GluAspMetLeuLeuSerHisLysProLysVal 340
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1103 ACTCAGATAGCGG-----CCTTCCTCATGCATCTCCTGGGAGTGGGAGGCCACA 1156
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 341 GluMetAspProArgArgPheProTyrTyrMetThrAspGluValValGluAlaAla 360
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1157 TGCCTGCTCCGCTGCTCAGGAGAGAGAGGCC-----GAGAAAGAA 1198
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 ThrMetCysMetValAlaGlnAlaGluGluAlaLeuAsnTyrGluLysValGlnThrGlu 380
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1199 CACTGCTCCAAAGTCCGTCGAGCAGATGATCTCGGAGGAATTTGGAAGTGTCTATCA 1258
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 381 AspGluLysLeuIleAsnMetGluLysLeuValLeuArgGluPheGlyArgCysLeuGlu 400
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1259 CAGATCTC 1267
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 401 GlnMetIle 403
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 2
B84585
hypoetical protein At2g20110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84585
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umavam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Cross-references: GB:AE002093; NID:94580462; PIDN:AAD24386.1; GSPDB:GN00139
A:Gene: At2q20110
A:Map position: 2
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <STO>
Alignment Scores:
Pred. No.: 5,47e-23 Length: 571
Score: 376.00 Matches: 95
Percent Similarity: 44.27% Conservative: 48
Best Local Similarity: 29.41% Mismatches: 90
Query Match: 9.83% Indels: 90
DB: 2 Gaps: 10
US-09-743-237-3 (1-2134) x B84585 (1-571)
```

Qy 527 TCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTTCAGGAGCATCAACAAA 586

```
Db 51 AlaSerIleProSerProIleValThrValThrArgProIleIleThrSerGlnAlaPro 70
Qy 587 TTAATCTCATTA-----CAACAACCTTGAGGAGCCTTA----- 622
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 ProThrValIleThrProIleProProGlnSerGlnGlyIleIleLeuHisVal 90
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 623 -----CCATCGGTAGTCAACGGGTCTGCTTTCCCTCGGATCAACT 664
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 91 ProfileArgHisProArgProGluSerProAsnSerMetProArgProAlaGlyGluThr 110
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 665 CTTCAGAGACACCAAAA-----ATAACTTTGGCT 694
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 111 ArgAspGlyThrProGlnLysLysLysGlnCysAsnCysLysHisSerArgCysLeuLys 130
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 695 GGGTACTGTGACTGCTTTGGCAGTGGGAGCTTTTGCACCAACTGCAATTTGTAATAATTGT 754
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 131 LeuTyrCysGluCysPheAlaSerGlyThrTyrCysaspGlyCysAsnCysValAsnCys 150
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 755 TGCACAACCTTGCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGTCTTGGTAGA 814
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 151 PheAsnAsnValGluAsnGluProAlaArgGlnAlaValGluSerThrLeuGluArg 170
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 815 AATCCACAAGCTTTTCCAGCCCAAAATTT-----GGGAAG 847
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 171 AsnProAsnAlaPheArgProLysIleAlaAlaSerProHisGlyGlyArgAspAsnArg 190
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 848 GGCAATTTGGGCAATGTC-----AAGCCCCAGCACCAACAAAGGTGCAACTGCAGGAGG 901
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 GluGluValGlyAspValMetLeuAlaArgHisAsnLysGlyCysHisCysLysLys 210
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 902 TCAGGCTGCTGAAGAATTTACTGCGAGTGTATGAGGCCCAAAATTTGTTCTTCTATT 961
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 211 SerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIleLeuCysSerGluAsn 230
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 962 TGCAAAATGCAATGCTTCCAAAATTTATGAAGAAAGCCAGAACCAAGACACATAATGAGC 1021
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 231 CysLysCysAspCysLysAsnPheGluCysLeuValArgGlnSerLeuPheHis 250
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1022 ATGCCAAAC-----TACATGCAG-----ACTGGAGGT 1048
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 251 GlyGluHisSerHisAsnLeuAlaTyrLeuGlnHisAlaAsnAlaAlaIleThrGlyAla 270
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1049 TTGAAGGCGAGCCATTACCTG-----CCACCACGAATAATTTTCAGGA----- 1090
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 271 IleGlySerSerGlyPheAlaSerAlaProProLysArgArgGlyGlnGluIle 290
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1091 -----CTTCCAGATTCAGTCACGAT 1111
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 291 PhePheAsnGlnGlyThrLysAspSerSerThrHisArgLeuGlyGlnAlaAsnAsnGly 310
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1112 AGCGGCGCTTCTCTCA-----TGATCTCTCTGGGAGGTGGTGGAGGCCACA 1156
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 311 ArgThrThrSerSerGlnThrGlySerArgAlaGlyGlyAsnAlaSerLeuGlyProSer 330
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1127 -----TGATCTCTCTGGGAGGTGGTGGAGGCCACA 1156
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 331 LysValValTyrLysSerLeuLeuAlaAsnIleIleLysPrometAspValLysAlaLeu 350
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1157 TGCCTGCTGCTGCTGCTGAGGAGAGAGGCCCGAGAGAAACACTGCTCCAACTGCTG 1216
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 351 CysSerValLeuValAlaValAlaGlyGluAlaLysThrLeuThrGluLysArgLeu 370
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1217 GCAGAGCAG 1225
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 371 AlaAsnGln 373
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 3
T08955
hypoetical protein F19B15.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08955
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
```

Db 290 PheAlaProSerProAlaProLysArgArgLysGlyGlnGluIleLeuPheAsnGlnAla 309

RESULT 4

F71410
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C:Accession: F71410
R:Bavan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.;
P.: Wedler, H.; Wedler, E.; Wambutt, R.; Weltzenegger, T.; Pohl, T.M.; Terry, N.;
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk,
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chlawatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: F71410
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-658 <BEV>
A:Cross-references: GB:297337; NID:g2244829; PID:e326824; PID:g2244834
C:Genetics:
A:Map position: 4COP9-4G3845

Alignment Scores:

Pred. No.:	2,43e-12	Length:	658
Score:	248.00	Matches:	92
Percent Similarity:	35.07%	Conservative:	49
Best Local Similarity:	22.89%	Mismatches:	143
Query Match:	6.49%	Indels:	118
DB:	2	Gaps:	14

US-09-743-237-3 (1-2134) x F71410 (1-658)

QY	287	AATATCGGAAGAACAACTTTGCAGAACTTCTTGCTCAGCAATCCTGTTCGAAGTTC	346
Db <td>219<td>AsnGlyValGluLysGlnThrMetGlnHis-----</td><td>228</td></td>	219 <td>AsnGlyValGluLysGlnThrMetGlnHis-----</td> <td>228</td>	AsnGlyValGluLysGlnThrMetGlnHis-----	228
QY <td>347<td>CCATGGTCCCAGGAACTAGAGGATGCCCTGCTGCTTCTTAAGAAGATTCCAACCCA</td><td>406</td></td>	347 <td>CCATGGTCCCAGGAACTAGAGGATGCCCTGCTGCTTCTTAAGAAGATTCCAACCCA</td> <td>406</td>	CCATGGTCCCAGGAACTAGAGGATGCCCTGCTGCTTCTTAAGAAGATTCCAACCCA	406
Db <td>229<td>-----AspSerAsnLysGluProGluSerAlaAsnAlaIleProTyrGluValasr</td><td>246</td></td>	229 <td>-----AspSerAsnLysGluProGluSerAlaAsnAlaIleProTyrGluValasr</td> <td>246</td>	-----AspSerAsnLysGluProGluSerAlaAsnAlaIleProTyrGluValasr	246
QY <td>407<td>ATGGTGATATGCCAA-----TTGAAGGGGGGCACACAATGCTATGTATAGAC</td><td>454</td></td>	407 <td>ATGGTGATATGCCAA-----TTGAAGGGGGGCACACAATGCTATGTATAGAC</td> <td>454</td>	ATGGTGATATGCCAA-----TTGAAGGGGGGCACACAATGCTATGTATAGAC	454
Db <td>247<td>GlyValIleSerGlnAlavalserLeuLeuHisAtgGlylleArgArgCysLeuasp</td><td>266</td></td>	247 <td>GlyValIleSerGlnAlavalserLeuLeuHisAtgGlylleArgArgCysLeuasp</td> <td>266</td>	GlyValIleSerGlnAlavalserLeuLeuHisAtgGlylleArgArgCysLeuasp	266
QY <td>455<td>-----AANTCTAGACACAGACTA</td><td>475</td></td>	455 <td>-----AANTCTAGACACAGACTA</td> <td>475</td>	-----AANTCTAGACACAGACTA	475
Db <td>267<td>PheGluMetProGlyAsnLysGlnThrSerSerGluAsnThrAlaLaCysGluSer</td><td>286</td></td>	267 <td>PheGluMetProGlyAsnLysGlnThrSerSerGluAsnThrAlaLaCysGluSer</td> <td>286</td>	PheGluMetProGlyAsnLysGlnThrSerSerGluAsnThrAlaLaCysGluSer	286
QY <td>476<td>AAAGCATCCATTGTTGTTCTCAGTAT-----</td><td>502</td></td>	476 <td>AAAGCATCCATTGTTGTTCTCAGTAT-----</td> <td>502</td>	AAAGCATCCATTGTTGTTCTCAGTAT-----	502
Db <td>287<td>SerSerArgCysValIleProSerIleGlyLeuHisLeuAsnAlaIleLeuMetSer</td><td>306</td></td>	287 <td>SerSerArgCysValIleProSerIleGlyLeuHisLeuAsnAlaIleLeuMetSer</td> <td>306</td>	SerSerArgCysValIleProSerIleGlyLeuHisLeuAsnAlaIleLeuMetSer	306
QY <td>503<td>CAAGATCAAATAATTATCTACAGTCAGATGTCCTTAAACCAGTACTGCTTTAGTAGGG</td><td>562</td></td>	503 <td>CAAGATCAAATAATTATCTACAGTCAGATGTCCTTAAACCAGTACTGCTTTAGTAGGG</td> <td>562</td>	CAAGATCAAATAATTATCTACAGTCAGATGTCCTTAAACCAGTACTGCTTTAGTAGGG	562
Db <td>307<td>LysAspCysLysThrAsnValThrGlnAspTyrSerCysSerAlaAsnIleGlnValGly</td><td>326</td></td>	307 <td>LysAspCysLysThrAsnValThrGlnAspTyrSerCysSerAlaAsnIleGlnValGly</td> <td>326</td>	LysAspCysLysThrAsnValThrGlnAspTyrSerCysSerAlaAsnIleGlnValGly	326
QY <td>563<td>AGATTTTGGCAGCATCAACAANAATAATCTTCATTACACAACTTGGAGGAGCCTTA</td><td>622</td></td>	563 <td>AGATTTTGGCAGCATCAACAANAATAATCTTCATTACACAACTTGGAGGAGCCTTA</td> <td>622</td>	AGATTTTGGCAGCATCAACAANAATAATCTTCATTACACAACTTGGAGGAGCCTTA	622

Db 327 LeuGlnArgSerIleSerThr-----LeuGlnAspSerLeu 338
QY 623 CCATCGGTAGTCAACGGTCTGCTTCCCTCGGATCACTCTCCAGGACCACCA--- 679
Db 339 AspGlnThrGluAsnGluIleArgGluAspAlaAspGlnAspValProAla 358
QY 680 -----AAATAACCTTTGGCT-----GGTACTGTGACTGCTTTGCCAGT 718
Db 359 LeuGlnGluAsnLeuSerSerProIlyLysLysSerTyrcysGluCysPheAlaAla 378
QY 719 GGGGACTTTTGC---AACAACTGCAATGTAATAATTTGCAACAACCTTCATCATCAT 775
Db 379 GlyValTyrcysIleGluProCysSerCysIleAspCysPheAsnLysProIleHisGlu 398
QY 776 ATTGAACGGTTTAAAGCATTAAGCGATGCTCTGTGAGAAATCCAGAGCTTTCCAGCCA 835
Db 399 AspValValLeuAlaThrArgGlyGlnIleGluSerArgAsnProLeuAlaPheAlaPro 418
QY 836 AAAATTGGGAAGGCG-----CAATTGGGCAAT----- 862
Db 419 LysValIleArgAsnSerAspSerValGlnGluThrGlyAspAspAlaSerLysThrPro 438
QY 863 GTCAGCCCGACACACAAAGGTCAGCTGAGGAGGTCAGGCTGCCTGAAGAATTAC 922
Db 439 AlaSerAlaArgHisLysArgGlyCysAsnCysLysLysSerAsnCysLeuLysLysTyr 458
QY 923 TGCAGGTCTATAGGCCCAATATGTTCTTCTATTGCAATTCATGCTGGTGCATA 982
Db 459 CysGluCysTyrGlnGlyValGlyCysSerIleAsnCysArgCysGluGlyCysLys 478
QY 983 AAT-----TATGAAGAAGC 997
Db 479 AsnAlaPheGlyArgLysAspGlySerSerIleAspMetGluAlaGluGlnGluGlu 498
QY 998 CCAGAA----- 1003
Db 499 AsnGluThrSerGluLysSerArgThrAlaLysSerGlnGlnAsnThrGluValLeuMet 518
QY 1004 CGAAGACACTAATGAC-----ATGCCAACTAC 1033
Db 519 ArgLysAspMetSerSerAlaLeuProThrThrProThrProIleTyArgProGluLeu 538
QY 1034 ATGCAGACTGGAGGTTTGGAGGCAGCATTACCTGCCACCAACAGAA----- 1081
Db 539 ValGlnLeuProPheSerSerLysAsnArgMetProProGlnSerLeuLeuGly 558
QY 1082 -----TTTCAGACTTCCAGATTCAAGTCAAGTACGATGCGGCTTCTCATCATC 1132
Db 559 GlyGlySerSerGlyIlePheAsnSerGlnTyLeuArgLysProAspIleSerLeu 578
QY 1133 TCCTGGGAGGTGTGGAGGCCACATGCGCTGCTGCTGCTGAGGAGAGAGCGCGAG 1192
Db 579 SerGlnSerArgIleGluLysSerPheGluThrValAlaValAspGlyAlaGluGlnMet 598
QY 1193 AAAGAA 1198
Db 599 ProGlu 600
RESULT 5
S08167
Balbiani ring 3 protein - midge (Chironomus tentans)
C:Species: Chironomus tentans
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 21-Jul-2000
C:Accession: S08167
R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.
J. Mol. Biol. 211, 331-349, 1990
A:Title: The balbiani ring 3 gene in Chironomus tentans has a diverged repetitive structure
A:Reference number: S08167; MUID:90172404; PMID:1689777
A:Accession: S08167
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1700 <PAU>
A:Cross-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058

C:Genetics:
A:Gene: BR3
A:Map position: 4
C:Superfamily: unassigned Balbiani ring proteins

Alignment Scores:
Pred. No.: 0.0618 Length: 1700
Score: 123.00 Matches: 87
Percent Similarity: 31.73% Conservative: 45
Best Local Similarity: 20.91% Mismatches: 146
Query Match: 3.22% Indels: 139
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x S08167 (1-1700)

QY 701 TGTGACTGCTTTGCCAGTGGGAG-----TTTTCG---AAC 733
Db 198 CysSerCysGluCysLysGlyAspGlyLysCysGlnGlySerLysIleTyrCysLysAsn 217
QY 734 AACTGCAATTGT-----AATAATTGTTGCAACAACCTTCAT 769
Db 218 AsnCysArgCysIleCysProThrAlaGluProAlaGlyGlyCysSerAlaProLeuLys 237
QY 770 CATCATATTCAACGGTTTAAAGCCATTAAAGCATCTCTGTGAGAAATCCAGAGCTTTC 829
Db 238 TrpAspAspAsp-----LysCysSerCysAlaCysProAlaLysMetGluGluLysLys 255
QY 830 CAGCCAAAAATTGGGAAGGCGCAATTGGCAATGTCAAGCCCGACGACCAACAAGGGTGC 889
Db 256 GluLysCysValGluSerGlyLysIleTyrPasn-----ProAsnThrCysGluCys 272
QY 890 AACTGCGAGGAGTGCCTGCCTCAAGAAT-----TACTGC 925
Db 273 GlyCysAlaGlnLeuAsnCysProAspAsnLysLysAlaAsnLysGluThrCysGlnCys 292
QY 926 GACTGCTATGAC-----GCCCAATTATGTGTTCTTCTATTTCGAAA 967
Db 293 GluCysLysGluValLysLysCysAsnGlyGlyGlnValPheCysLysAspSerCysSer 312
QY 968 TGCAATTGTTGCCAAAAATTATGAAGAA-----AGCCAGAACCAAGACACA 1012
Db 313 CysVal---CysProGlyGlyAspLysAspLysThrCysThrAlaProGlnValTyArg 331
QY 1013 CTAATGAGCATGCCAACTACATCAGACTGGAGGTTTGGAGGCGACCATACCTACGCCA 1072
Db 332 GlyVal-AlaCysSerCysSerCysProValAsnMetGlnLysProAlaAspGlyCysPr 351
QY 1073 CCAACGAAATTTTCAGGACTTCCAACTTC-----AGTCACCATAG 1113
Db 351 oArgProGlnLysTrpAspLysGluGluCysArgCysGluCysProValLysHisAspCy 371
QY 1114 GGGGCTTCTCATGTCATCTCTGCGGAGGTGGTGGAGGCCACATGCGCTGCTGCTGCTGC 1173
Db 371 sLysAsnGlyLysVal-----TrpAsp-----GluThrIleCysGlnCysPr 387
QY 1174 TCAG-----GGAGAAGAGCGCGAG----- 1192
Db 387 oArgAspAlaProValCysThrAlaGlyLysGluArgCysGlyGluSerCysGluCysLy 407
QY 1193 -----AAAGACACTGCTCCCAAG----- 1210
Db 407 sCysIleAsnArgGluProLysGluGlyCysAlaLysProLeuValTrpAsnGluAsnTh 427
QY 1211 -----TCCTGGCAGACAGATGATCTCG-----GA 1236
Db 427 rCysLysCysValCysProAlaAspLysGlnMetSerProGlyGlyCysGlySerGlyLy 447
QY 1237 GGAATTTGGAAGTGCCTTATACAGATTTCCACACTGAGTTTAAATCTTAAG---GGATT 1293
Db 447 sSerPheAsnLysLeuThrCysGlnCysGluCysAspGlnSerAlaSerLysCysGlyLe 467
QY 1294 GAAATG-----GAGTAGAGTATTAAGTGTGAATGCTGTTGATTGTTGCTTCTAGCTAG 1347
|||||

```
Db 467 uLysArgTrpAsnAlaAAspThrCysLysCysGluCys----- 479
QY 1348 AAATCTCTAGTTTAAAGAGGATGTTTAGGGGAACATGAGGCTGGCTCTGCACCAACACC 1407
Db 480 -----GlnPr 481
QY 1408 AGGCTCCCTGCATCCCTGGGGCCAGGGAGTTTACTCA----- 1445
Db 481 oGlyMetProGlu---GlyCysGlyLys-GlnThrTrpLysSerAspLysCysLysC 500
QY 1446 --GAGCTCTCAAGATGGGCAACCCATGCCCCCTTTCTGAGGAGTGCATGGCCCTGA 1503
Db 500 ySGLysSerProThrIleThrCysGlnAlaProGlnIleLeuAspLeuAsnThrCysG 520
QY 1504 GCATTGTTTGT-----CTGGCCACAGAGAGAGCTGGGTTCCCATAGTCTCT 1551
Db 520 lUCysLysCysProValAsnMetLeuAlaGlnLysGlnLysCysLysSerProArgGlnT 540
QY 1552 GG-----GAGAGTGTCTGCAGGGGGGGGAGGAGGAGAGAGGAGGCT 1602
Db 540 rPThrAspSerLysCysLeuCysGluCysSerThrThrProAlaThrCysGluGlyLysG 560
QY 1603 CACTCTGGTCTGACTCTCTCTCAGAGATGTTGCTCTGGAGGC 1646
Db 560 lNThrTrpCysGlyGluAlaCysGlnCysIleCysProGlyGly 574

RESULT 6
MGF8 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00209
R:Nakajima, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T00209
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1737 <NAK>
A:Cross-references: EMBL:AB011541; NID:g3449307; PIDN:BAA32469.1; PID:g3449308
A:Experimental source: Brain; clone HGL392
C:Genetics:
A:Gene: MEGF8
A:Map position: 19q12

Alignment Scores:
Pred. No.: 0.0907 Length: 1737
Score: 121.00 Matches: 125
Percent Similarity: 29.04% Conservative: 60
Best Local Similarity: 19.62% Mismatches: 187
Query Match: 3.16% Indels: 265
DB: 2 Gaps: 35

US-09-743-237-3 (1-2134) x T00209 (1-1737)
QY 11 CAAGCGAAGCTCGCGGGGGGCGACAGCGCGGGAGCTCTCGGGAGTACCCCGG 70
Db 614 GlnGlyAlaLysArgAspArgMetArgAsnValArgGlySerSerArgGlyLeuGln 533
QY 71 GATCCAGAGCTCAGCGCTCGGAGAGCGTCCGCTCTGAGCGCCCGCCGCGCCCGC 130
Db 634 ValPro-----GlyGluGlnProGlySerTrpGlyPheArgGluValArg 648
QY 131 CTGCAACGTGCACTT---CCTGCTCTGCTGCTATCCCGCGCACCGCGCGCGGTGT 187
Db 649 LysLysMetAlaLeuTrpAlaLeuAlaGlyThr-----GlyGlyPhe 663
QY 188 TTGCCCCCTGGCGCTGGGTCTCT---GCGAAGGAGCGCTCCACCGCGCGCGCATGA 244
Db 664 LeuGlu-----GluIleSerProHisLeuLysGluProArgProArgLeuPheHisAla 681
QY 245 TCCCAAGTTAAATCAAGGTAGCAGGTGGTACTACTACA-----AGTAATAT 292
```

```
Db 682 SerAlaLeu-----LeuGlyAspThrMetValValLeuLeuGlyGlyArgSerAsp 697
QY 293 CCGGAAGAAGCAACTTTTCAGAAATCTTTCGTCAAGAAATCTGTTCAGATTCATCCATGG 352
Db 698 ProAspGluPheSerSerAspValLeuLeuTyrglnValAsnCys----- 712
QY 353 TCCCAGGAACTAGAGGATGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 412
Db 713 -----AsnAlaTrpLeu 716
QY 413 ATATGCCAATTTGAAGGGGGCACACAAATGCTATGTATAGACAATTTCTAGACAAGAA 472
Db 717 LeuProAspLeu-----ThrArg--- 722
QY 473 CTAAAAGCACTCCATTTGGTTCTCTCAGTATCAAGATCAAAATATTTATCTACATCAGAT 532
Db 723 -----SerAlaSer 725
QY 533 GTCCCTAAACCAATGACTGCTTTTAGTGGGAGATTTTTCGCCAGCA--TCACAAAATTA 589
Db 726 ValGlyProProMetGluSerValAlaHisAlaValAlaValGlySerArgLeu 745
QY 590 AATCTATTACACAACAACTTGGAGGAGCTTTACCATCGGTAGTCAACGGGTCTGTCTTC 649
Db 746 -----TyrIleSerGlyGlyPheGlyValAlaLeuGlyArgLeuLeu 760
QY 650 CCTCGGGATCAACTCTTCCAGGAGCACCAAAATAACTTTGGCTGGGTACTGTGACTGC 709
Db 761 -----AlaLeuThrLeuProAspPro-----CysArgLeu 771
QY 710 TTTGCCAGTGGGACTTTTTCACAACAC--TGCAAT--TGTAATAATTTGCTGTC 757
Db 772 LeuSerSerProGluAlaCysAsnGlnSerGlyAlaCysThrTrpCysHisGlyAlaCys 791
QY 758 -----AACAACTTGCAT----- 769
Db 792 LeuSerGlyAspGlnAlaHisArgLeuGlyCysGlySerProCysSerProMetPro 811
QY 770 -----CATCATATTGAACGGTTTAAGCCATTAAGCCATGCTTTCGTAGAATCCA 820
Db 812 ArgSerProGluGluCysArgArgLeuArgThrCysSerGluCysLeuAlaArgHisPro 831
QY 821 GAAGCTTTCCAGCAAAATTTGGAGGAGGCAATTTGGGCAATGTCAAGCCCGCAGCAAC 880
Db 832 ArgThrLeuGlnPro-----GlyAspGlyGluAlaSerThr-----ProArgCysLys 847
QY 881 AAAGGTGCAACTGCAGGAGTCAAGCTGCCTG-----AAG 916
Db 848 TrpCysThrAsnCysProGluGlyAlaCysIleGlyArgAsnGlySerCysThrSerGlu 867
QY 917 AATTACTCGAGTGTATGAGGCCCAAAATATGTGTTCTTCTTCTATTTC----- 964
Db 868 AsnAspCysArgIleAsnGlnArgGluValPheTrpAlaGlyAsnCysSerGluAlaAla 887
QY 965 -----AAATGCATTGGTTGCAAAAT 985
Db 888 CysGlyAlaAlaAspCysGluGlnCysThrArgGluGlyCysMetTrpThrArgGln 907
QY 986 TATGAAGAAGCCCGAGACGAAGACACTTAATGAGCATG---CCAAACTAC----- 1033
Db 908 PheLysArgThrGlyGluThrArgArgIleLeuSerValGlnProThrTrpArgTrpThr 927
QY 1034 -----ATGCAGACTGGAGTTTGAAGCGCAGCATTAACCTG 1069
Db 928 CysPheSerHisSerLeuLeuAsnValSerProMetProValGluSerSerProLeu 947
QY 1070 CCA---CCAACCAAAATTTTCAGGAGCTTCCAAGATTTCAGTCAAGATAGGCGGCTTCTCTCA 1126
Db 948 ProCysProThrProCysHisLeuLeuPro-----Asn 958
QY 1127 TGCATCTCTCGGAGGTGGTGGAGGCCACATCGCTCTGCTGCTCTGCTCAGGAGAGAG 1186
Db ----- 1199
```

```

Db 959 CysThrSer-----CysLeuAspSerLysGlyAlaAsp 969
Qy 1187 GCCGAGAAAGAACACTGCTCAAGTGCTGGCAGAGCAGATGATCCTGGAGAAATTTGGA 1246
Db 970 GlyGlyTrpGlnHisCys-----ValTrpSer 978
Qy 1247 AGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGATTGAAATGGAGTAG 1306
Db 979 SerSerLeuGlnCysLeuSerProSerTyr----- 989
Qy 1307 AGTATAAGTGTGATGCATGCTGATTTGCTTTAGTCTAGATAATCTAGTTTAGAAG 1366
Db 990 ---LeuProLeuArgCysmet----- 995
Qy 1367 GATGTTTAGGGGAACATGAGGCTGGCTGTCAGCAACAACAGGCTCCCTGTCATCCCTG 1426
Db 996 -----AlaGly 997
Qy 1427 GGCCGAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCATGCCCCCTTTTCTGA 1486
Db 998 GlyCysGlyArgLeuLeuArgGly----- 1005
Qy 1487 GGAGTGCTGCTGAGCATTGTTGTCGGCCAGAGGAGAGCTTGGGTTCCCATTA 1546
Db 1006 -----ProGluSerCysSerLeuGlyCysAlaGlnAlaThrGlnCysAlaLeu 1021
Qy 1547 GTCCTG-----GGAGAGTGCTGCGAGGCGGC----- 1573
Db 1022 CysLeuArgArgProHisCysGlyTrpCysAlaTrpGlyGlnAspGlyGlyArg 1041
Qy 1574 -----GGAGGCGAGCAGGCCCTCGCGAG-----AGCTC 1603
Db 1042 CysMetGluGlyGlyLeuSer-GlyProArgAspGlyLeuThrCysGlyArgProGlyAl 1061
Qy 1604 ACTCTGGTCGACTCT-----CCTCTCAGAGATGTTGCTCTGGA 1643
Db 1061 aSerTrpAlaPheLeuSerCysProProGluAspGluCysAlaAsnGly 1077

RESULT 7
T00532
Probable cadmium-transporting ATPase (EC 3.6.1.-) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999
C:Accession: T00532
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1997
A:Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A:Reference number: Z14167
A:Accession: T00532
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1001 <ROU>
A:Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176713
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 27/3; 113/1; 159/1; 268/2; 336/1; 403/3; 980/1
A:Note: T20K24.13
C:Superfamily: ATPase nucleotide-binding domain homology
C:Keywords: hydrolase
F:360-504/Domain: ATPase nucleotide-binding domain homology <ATN>

Alignment Scores:
Pred. No.: 0.119 Length: 1001
Score: 119.50 Matches: 110
Percent Similarity: 29.72% Conservative: 57
Best Local Similarity: 19.57% Mismatches: 184
Query Match: 3.12% Indels: 211
DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x T00532 (1-1001)
Qy 235 GTCCGCATGATCCCGAGTTGAATCAAGGTAGCAGGTGGTACTACTACAAGTAATATCC 294

```

```

Db 476 IleArgIleProGlnAlaValLysLeuAlaArg-----Arg 488
Qy 295 GGAAGAAGCAACTTTGCGAGAATCTTCTGCTCAGGAATCTGTTGCAAGTTCCCATGTC 354
Db 489 AlaArgIysValValGluAsnValCysLeuSerIleIleLeu-Lys----- 504
Qy 355 CCAGGAACCTAGAGGATGCCTCCTGCTGCTCTTCTTAAAGAAATTCACCAACCAATGGTG-- 412
Db 505 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProLeuIleTr 519
Qy 413 -----ATATGCCAATTGAAGGGGCACAAATGCTATGTATAGACAACTCTAGA----- 463
Db 519 pAlaAlaValLeuValAspValGlyThrCysLeuValIlePheAsnSerMetLeuLe 539
Qy 464 -ACAAGAGAAGCTAAAGCACTCCATTTGGTTCCTCAGTATCAAGATCAAAATATATTATCT 522
Db 539 uLeuArgGluLysLysIleGlyAsnLysLysCysTyrArg----- 553
Qy 523 ACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAAC 582
Db 554 -----AlaSerThrSe 557
Qy 583 AAAATTAATCTCATTACAAACAACTTGAGGGAGCCTTACCATCGGTAGTCAAC----- 637
Db 557 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyrValValAspLeuGl 575
Qy 638 ---GGGTGCTGCTTCCCTCGGATCACTCTCCAGGACCACCAAAATAAATTTTGGC 693
Db 575 uAlaGlyLeuLeuThrLysSerGlyAsn----- 584
Qy 694 TGGTACTGTGACTGCTTGGCAGTGGGACTTTTGCAACAACATGCAAT----- 742
Db 585 -GlyGlnCysLysSerSerCysCysGlyAspLysLysAsnGlnGluAsnValValMetMe 604
Qy 743 -----TGT----- 745
Db 604 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysGlyAspLysLy 624
Qy 746 -----AATAATTGTTGC----- 757
Db 624 sGluGluLysValLysProLeuValLysAspGlyCysSerGluLysThrArgLysSe 644
Qy 758 -----AACAACTGTCATCATGATAT 777
Db 644 rGluGlyAspMetValSerLeuSerSerCysLysLysSerHisValLysHisAspLe 664
Qy 778 TGAACGGTTTAAAGCCATTAGGCATGCTTGGTAGAAATCCAGAACTTCCAGCCAAA 837
Db 664 u---LysMetLysGlyGlySerGlyCysCysAlaSerLysAsnGluLysGlyLysGluVa 683
Qy 838 AATTGGGAAGGCCCAATTGGCAATGTCAAGCCCCCAGCACAAAGGGTGCACACTGCAG 897
Db 683 lValAlaLysSerCysCysGluLysProLysGlnGlnValGluSerValGlyAspCysLy 703
Qy 898 GAGTCAGGCTGCCTGAAG----- 916
Db 703 sSerGlyHisCysGluLysLysGlnAlaGluAspIleValProValGlnIleIl 723
Qy 917 -----AATTACTGGAGTGCTATGAGGCCCAAAATATGTTCTTCTATTTCGAATGCAT 972
Db 723 eGlyHisAlaLeuThrHisValGluIleGluLeuGlnThrLysGluThrCysTyrThrSe 743
Qy 973 TGGTTGCAAAAATTTATGAAGAAAGCCAGAAAGACATATAGCATGCCAAAC-- 1030
Db 743 rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuSerSerGluAsnTh 763
Qy 1031 ----TACATGCAGACTGGAGTTTG-----GAAGCAAGCA 1062
Db 763 rProTyrLeuGluLysGlyValLeuIleLysAspGluGlyAsnCysLysSerGlySerGl 783
Qy 1063 TTACCTCCCAACCAAGAAATTTTCAGGACTTCCAAGATTCAAGTTCAGTCACCATAGGCGCCTTC 1122

```

Db 783 uAnMetGlyThrValylsGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh 803
 QY 1123 CTCATCATCTCTCTGGAGGTGGTGGAGGCCACATGGCTGCTGCTCAGGAGA 1182
 Db 803 rGlyGluLeuThr-----LeuAlaSerGluG1 812
 QY 1183 AGAGCGCGAGAAGAACTGCTCCAG-----TCCTGGCAGAGACAGATGATCTCGGA 1236
 Db 812 uGluThrAspGlnAspCysSerGlyCysValAsnGluGlyThrValLysG1 832
 QY 1237 GGAATTGGAGGTCTTATCATCAGATTCTCCACATGATTTAAATCTAAGGATGAA 1296
 Db 832 nSerPheAspGluLysHisSerValLeu-----ValGluLysGluGlyLeuAs 849
 QY 1297 AATGGAGTAGATATAAGTGTCAATGCTGTTGATTTGCTCTAGCTAGAAATCTCTA 1356
 Db 849 pMetGluThrGlyPheCysAspAlaLysLeuValCys----- 862
 QY 1357 GTTTAGAAAGGATTTAGGGGAACATGAGGCTGCTGCAGCAACAACAGGCTCCCC 1416
 Db 862 ----- 862
 QY 1417 TGCATCCCTGGCCAGGGAGTTTACTCAGAGCTCTCTGAGATGTGGCAACCCATGCC 1476
 Db 863 ----- 866
 QY 1477 CCTTTCTGAGAGGTGCATGCGCTGAGCATTTGTCTGGCCAGAGGAGAGAGC 1533
 Db 867 -----GluGlyGluValLysGlnCysArgLeuGluLeuLysLysGluGlu-- 882
 QY 1534 TTGGTGTCCCATAGTCTCTGGAGAGTGTCTGCAGGCGCGGAGGAGCAGAGGCCCTG 1593
 Db 883 -----HisCysLysSerGlyCysGlyCysGlyGluLeuGlnTh 895
 QY 1594 CGGAGAGCTCACTCTGTCGACTCTCTCTCTCAGCAAGT-----TGCTCTGGAGGCT 1647
 Db 895 rGlyGluLeuThrLeuValSerGluGluGlu-ThrGluSerThrAsnCysSerThrGlyC 915
 QY 1648 GC 1649
 Db 915 ys 915
 RESULT 8
 F84572
 probable cadmium-transporting-ATPase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84572
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: AB4420; MUID:20083487; PMID:10617197
 A:Accession: F84572
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1172 <STO>
 A:Cross-references: GB:AE002093; NID:94210504; PIDN:ADI2041.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g19110
 A:Map position: 2

Alignment Scores:
 Pred. No.: 0.12 Length: 1172
 Score: 119.50 Matches: 110
 Percent Similarity: 29.72% Conservative: 57
 Best Local Similarity: 19.57% Mismatches: 184
 Query Match: 3.12% Indels: 211
 DB: 2 Gaps: 23

US-09-743-237-3 (1-2134) x F84572 (1-1172)

QY 235 GTCCGCATGATCCCATGTTGAAATCAAGGTAAAGCAGGTGGTACTTACAAAGTAATAATCC 294
 Db 639 IleArgIleProGlnAlaValLysLeuAlaArg-----Arg 651
 QY 295 GGAAGAAGCAACTTTCAGAAATCTTCTGCTCAGGAATCTCTGTGCAAGTTCCCATGGTC 354
 Db 652 AlaArgArgLysValValGluAsnValCysLeuSerIleIleLeu-Lys----- 667
 QY 355 CCAGGAACATAGAGGATGCTCTGCTGCTTCTTAAGAAAGATCCCAACCAATGGTG-- 412
 Db 668 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProLeuIleTr 682
 QY 413 ----ATATGCAATTGAAAGGGGCACACAAATGCTATGTATAGAAATCTCTAGA----- 463
 Db 682 pAlaAlaValLeuValAspValGlyThrCysLeuLeuValIlePheAsnSerMetLeuLe 702
 QY 464 -ACAAGAGAACTAAAGCACTCCATTTGGTTCTCTCAGTCAAGATCAAAATAATTATCT 522
 Db 702 uLeuArgGluLysLysIleGlyAsnLysLysCysTyrArg----- 716
 QY 523 ACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAAC 582
 Db 717 -----AlaSerThrSe 720
 QY 583 AAAATAAATCTCATACACAACTTGAGGAGCCTTACCATCGGTAGTCAAC----- 637
 Db 720 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyrValValAspLeuG1 738
 QY 638 ----GGTCTGCTTCCCTCGGGATCACTCTTCCAGGACCACCAAAATAACTTTGGC 693
 Db 738 uAlaGlyLeuLeuThrLysSerGlyAsn----- 747
 QY 694 TGGGTACTGTCAGTCTTCCAGTGGGACTTTTCCAACTGCAAT----- 742
 Db 748 -GlyGlnCysLysSerSerCysGlyAspLysLysAsnGlnGluAsnValValMetMe 767
 QY 743 -----TGT----- 745
 Db 767 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysGlyAspLysLys 787
 QY 746 -----AATAATTGTTGC----- 757
 Db 787 sGluGluLysValLysProLeuValLysAspGlyCysSerGluLysThrArgLysSe 807
 QY 758 -----AACAACCTGTCATCATGATAT 777
 Db 807 rGluGlyAspMetValSerLeuSerSerCysLysLysSerSerHisValLysHisAspLe 827
 QY 778 TGAACGGTTTAAAGCCATTAAAGCATGTCTTGGTAGAAATCCAGAGCTTCCAGGCCAA 837
 Db 827 u----LysMetLysGlyGlySerGlyCysCysAlaSerLysAsnGluLysGlyLysGluVa 846
 QY 838 AATTGGAGAGGCCAATTGGGCATGTCAGCCGCCAGCACAAAGGGTGCAACTGTCAG 897
 Db 846 lValAlaLysSerCysCysGluLysProLysGlnValGluSerValGlyAspCysLys 866
 QY 898 GAGGTCAGGCTGCTGCAAG----- 916
 Db 866 sSerGlyHisCysGluLysLysLysGlnAlaGluAspIleValproValGlnIleIl 886
 QY 917 ----AATTACTGCGAGTGTATGAGGCCCAAAATATGTTCTTCTTATTTGCAATATGCAT 972
 Db 886 eGlyHisAlaLeuThrHisValGluLeuGluLeuGlnThrLysGluThrCysLysThrSe 906
 QY 973 TGGTTCAAAAATATTGAAGAAAGCCAGCAACAAAGACACTAATGAGCATGCGCAAC-- 1030
 Db 906 rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuLeuSerSerGluAsnTh 926
 QY 1031 ----TACATGCCACTGGAGGTTTG-----GAAGGCAGCCA 1062
 Db 926 rProTyrLeuGluLysGlyValLeuLysAspGluGlyAsnCysLysSerGlySerG1 946
 QY 1063 TTACCTGCCACCAACGAAATTTTTCAGGACTTCCAAAGATTTCAGTCACGATAGGGCGCTTC 1122

Db 946 uasnMetGlyThrValLysGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh 966
QY 1123 CTCATGCATCTCTCGGAGGTGGTGGAGCCACATGCGCTGCTGCTCAGGGAGA 1182
Db 966 rGlyGluLeuThr-----LeuAlaSerGluGI 975
QY 1183 AGAGCGCGAGAAGAACTGCTCAAG-----TGCTGGCAGAGCAGATGATCCTGGA 1236
Db 975 uGluThrAspGlnAspCysSerSerGlyCysValasnGluGlyThrValLysGI 995
QY 1237 CGAATTTGGAAGTCTTATCATCAGATCTCCACACTGAGTTTAAATCTAAGGGATGAA 1296
Db 995 nSerPheAspGluLysLysHisSerValLeu-----ValGluLysGluGlyLeuas 1012
QY 1297 ATGAGGTAGACTATAAGTGTGAATGATGTTGATTTGCTTAGTCTAGAAATCTCTA 1356
Db 1012 pMetGluThrGlyPheCysCysAspAlaLysLeuValCys----- 1025
QY 1357 GTTTAGAAAGGATGTTTAGGGGAACATGAGGCTGCTCTGCAGACAACACGAGGCTCCCC 1416
Db 1025 ----- 1025
QY 1417 TGCATCCCTGGCCCGAGGAGTTTACTCAGAGCTCTCTCAAGATGTGGCAACCCATGCC 1476
Db 1026 -----CysGlyAsnThr----- 1029
QY 1477 CTTTTCTGAGAGGTGATGCCTGACGATGTGT---TTGCTGCCCCAGAGGAGAGC 1533
Db 1030 -----GluGlyGluValLysGluGlnCysargLeuGluLysLysGluGlu-- 1045
QY 1534 TTGGTTCOCATAGTCTGGGAGAGTGTCTGCAGGCGCGGAGGCGAGAGCAGGCCCTG 1593
Db 1046 -----HisCysLysSerGlyCysGlyGluGluLeuGlnTh 1058
QY 1594 CGAGAGCTCACTGTGCTGACTCTCTCTCAGAGAAATGT-----TGCTGTGAGGCT 1647
Db 1058 rGlyGluLeuThrLeuValSerGluGluGlu-ThrGluSerThrAsnCysSerThrGlyC 1078
QY 1648 GC 1649
Db 1078 ys 1078

RESULT 9

GNWVCJ
genome polyprotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C:Accession: A39253; PS0086
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotohno, K. Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A:Reference number: A39253; MUID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAAL4233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K. Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variations
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KA2>
A:Experimental source: Japanese isolate
A:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis B virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif A (P-loop)
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,224

Alignment Scores:

Pred. No.:	0.123	Length:	3010
Score:	119.50	Matches:	131
Percent Similarity:	32.74%	Conservative:	72
Best Local Similarity:	21.13%	Mismatches:	237
Query Match:	3.12%	Indels:	180
DB:	1	Gaps:	35

US-09-743-237-3 (1-2134) x GNWVCJ (1-3010)

QY 65 CCCGGGATCCAGAGCTCAGCGGCTGGAGACGTCGCGCTCGTCGAGGC----- 115
Db 1217 ProAlaValProGlnThrPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLys 1236
QY 116 -----CCGCGACCGCCCGCTGCAACGTGCACCTCTCTGCTCGCTGCTACC 163
Db 1237 SerThrLysValProAlaAlaThrAlaGln---GlyThrLysValLeuValLeuAsn 1255
QY 164 CGCGCACCGACCGCGGGTGTGTTGCGCCCTGGGCGCCCTG-----GGTCTCGC 214
Db 1256 -----ProSerValAlaAlaThrLeuGlyPheGlyAlaThrMetSerLysAlaHisGly 1273
QY 215 AAGAGCCTCCACCGCGGCGCTCGCATGATCCCATGTAATCAAGTACAGCAGTGGT 274
Db 1274 IleGluProAsnIleArgThrGlyVal-----Arg 1283
QY 275 ACTACTACAAGTAATAATCCGGAAGAACAACTTTGAGCAATCTTTGCTCAGGAATCC 334
Db 1284 ThrIleThrThrGlyGlyPro-----IleThrThrSerThr 1295
QY 335 TGTGCAAGTTCCTCCATGTCCTCCAGGAAGTACAGGATGCTCTGCTGTCTTAAAGAAA 394
Db 1296 TyrCysLysPhe-----LeuAlaAspGlyGlyCysSerGlyGlyAlaThr 1310
QY 395 GATTCCAACTTGGTATATGCCAATGAAAGGGGACACAAATCTATGTATA--- 451
Db 1311 Asp-----IleIleIleCysAspGluCysHisSerThrAspSerThrIleLeu 1327
QY 452 -----GACAATCTAGAACAGAGAACTAAAGAGCACTCCATTTGGTT 493
Db 1328 GlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAla 1347
QY 494 -----CCTCAGTATCAAGATCAAAATAAT 517
Db 1348 ThrAlaThrProProGlySerIleThrValProHisProAsnIleGluGluValAlaLeu 1367
QY 518 TATCTACAGTCAGATGCTCT-----AAACCAATCACTGCTTAGTAGGG 562
Db 1368 SerAsnThrGlyGluIleProPheThrGlyLysAlaIleProIleGluAlaIleLysGly 1387
QY 563 AGA-----TTTTTCCAGCATCAACAAATAATTAATCTCATACCAACACT 610
Db 1388 GlyArgHisLeuIlePheCysHisSerLysLysCysAspGluLeuAlaAlaLysLeu 1407
QY 611 GAGGGA---GCCTTACCATCCGTA-----GTCAAGGGTCTGCTTTCCCC 652
Db 1408 ThrGlyLeuGlyLeuAsnAlaValAlaThrArgGlyLeuAspValSerValIlePro 1427
QY 653 ---TCGGGATCACTCTCCAGGACCCACCAAAATAACTTTGGCTGGTACTGTGACTGC 709
Db 1428 ThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGlyPhe----- 1444


```
QY 710 TTGCGAGTGGGGACTTTTCCAAACAACAGTCAATTTGTAATAATTTGTCACAAACTTG--- 766
Db 1445 -----ThrGlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrVal 1462
QY 767 -----CATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGTCCT 808
Db 1463 AspPheSerLeuAspProThrPheThrIleGluThrThrThrLeuProGlnAspAlaVal 1482
QY 809 GGTAGAATCCAGAAAGCTTCCAGCCCAAAATTTGGGAAGGCCCAATTTGGGC----- 859
Db 1483 SerArgAlaGlnArg-----ArgGlyArgThrGlyArgGlyArgSerGlyIleTyrArg 1500
QY 860 AATGTCAGGCCCGCAGCACAAAGGTGCAACTGCAGGAGGTCCAGCTCCCTCAAGAAAT 919
Db 1501 PheValThrProGlyGlu-----ArgProSerGlyMetPheAspSer 1514
QY 920 -----TACTGCGAGTCTATGAGGCCCAATTTATGTGTTCTTCTATTTCCAATGCATT 973
Db 1515 SerValLeuCysGluCysTyrAspAla----- 1523
QY 974 GGTGCAAAATTTATGAAGAAAGCCAGAAAGACACTAATGAGCATGCCAAACTAC 1033
Db 1524 GlyCysAlaTrpThrGluLeuThrPro---AlaGluThrSerValArgLeuArgAlaTyr 1542
QY 1034 ATGCAAGACTGGAGTTTGGAA----- 1054
Db 1543 LeuAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluSerValPhe 1562
QY 1055 -----GCAGCCATTACTGCCACCACAGAAATTTTCAGGA----- 1090
Db 1563 ThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsn 1582
QY 1091 TTCCCAAGATTCAGTC-----ACGATAGGGGGCTTCCATCATCTCCTGG 1138
Db 1583 LeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProPr 1602
QY 1139 GAGGTGG-----TGGAGGCCACATGCGCCTGCTGCTCAGGGAAGAGAGGCC 1189
Db 1602 oSerTrpAspGlnMetTrpLys-----CysLeuIleArgLeuLysProTh 1617
QY 1190 GAGAAAGAACACTGCTCCCAAGTCCCTGGCAGACAGATGATCTCGGAGGAATTTGGAAG 1249
Db 1617 rLeuHisGlyProThrProLeuLeuTyrArgLeu-----G 1629
QY 1250 TGCTTATCACAGATTCTCCACACTGACT----- 1277
Db 1629 yAlaValGlnAsnGluValThrLeuThrHisProIleThrLysTyrIleMetAlaCysMe 1649
QY 1278 -----TTAAATCTAAGGATTGAAATGGAGTAGAGTATATAAGTGTGAATGCATG 1327
Db 1649 tSerAlaAspLeuGluValThrSerThrTrpValLeuValGlyValLeuAla 1669
QY 1328 TTGATTTTGTCTAGTCTAGAAATCTCTAGTTTGAAGAGATGTTTAGGGGAACATGAGG 1387
Db 1669 aLeuAlaAlaTyrCysLeuThrThrGlySerValIleVal-----G 1684
QY 1388 CTGCCTCTGCAGCAACAACAGCTCCCTCGCATCTCCCTGGCCCGAGGAGCTTTACTCAGA 1447
Db 1684 yArgIleIleLeuSerGlyArgProAlaValIleProAspArgGluValLeuThrGlnG 1704
QY 1448 GCTCTCT-----GAAGATGTGGCAACCCATGCCCTTTCTTGAGGAGGTGCATGCCCT 1501
Db 1704 uPheAspGluMetGluGluCysAlaSerHisLeuProTyrIleGluGln-----GlyMe 1722
QY 1502 GACCATTTGTTCTGCTGCCCGCAGAGAGAGAGCTTTGGTTCCTCCCATAGTCTCTGGGAGAGTGT 1561
Db 1722 tGlnLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeu-----LeuGlnThrAl 1739
QY 1562 CTCAGGCGCGCGGAGGCGCAGACAGCCCTCGCGAGAGCTACTCTGGTCGACTCTT 1619
Db 1739 aThrLysGlnAlaGluAla-----AlaAlaProValValGluSerLysTrpArgAlaLeu 1757
```

RESULT 10

T15881

hypothetical protein D1044_3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15881
R:Pauley, A.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of C. elegans cosmid D1044.
A:Reference number: Z18423
A:Accession: T15881
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1895 <PA>
A:Cross-references: EMBL:U00065; NID:g495681; PID:g495684; PIDN:AAA50735.1; CESP:D104
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:D1044.3
A:Introns: 35/2; 63/2; 150/1; 191/3; 207/2; 405/3; 551/3; 629/1; 764/1; 933/1; 959/2.

Alignment Scores:

Pred. No.:	0.133	Length:	1895
Score:	119.00	Matches:	107
Percent Similarity:	32.69%	Conservative:	80
Best Local Similarity:	18.71%	Mismatches:	178
Query Match:	3.11%	Indels:	207
DB:	2	Gaps:	30

US-09-743-237-3 (1-2134) x T15881 (1-1895)

```
QY 31 GCGACAGCAGCGGGGAGCTCTCGGGGAGTACCCCGGGATCCCGAGCTCAGCGCGC 90
Db 732 SerSerLeuAsnSerSerThrSerTyrLysArgProGlyArgSerSer-LysThrSe 751
QY 91 TGA---GGACGTGCGCTCTCGAGCGCCCGCAGCGCCCGCTGCAAGCTGACTTCC 147
Db 751 rGlySerGlyLysLeuProIleAsnProAlaProArgAsnHisSer----- 768
QY 148 TGCTCTGCTGCTACCGCGCACCGCGCGGGGTGTTTCCCTCGGGCGCTGGG 207
Db 769 ---MetSerAlaGlyGluThrProArgProSerSerIleValCysLeuProArgIleTh 787
QY 208 T-----CCTCGAAG----- 217
Db 787 rAspLysTyrSerValMetPheProSerAlaProSerAlaIleProSerArgArgVa 807
QY 218 -----GAGCTCCACCGCGCGCTCGCATGATCCAGTTGAAA---TC 258
Db 807 lGlnThrCysSerThrGluHisProAlaArgSerSerSerThrGluLeuLysValSe 827
QY 259 AAGSTAAGCAGGTGCTACTACTACAGTAATAAT----- 292
Db 827 rLysGlnSerAspGlyLeuThrValSerSerAsnValLeuArgGlnIleGlnAspProVa 847
QY 293 -----CCGGAAGAAGCAACTTTG-- 310
Db 847 lLeuThrIleLeuAsnGlnIleHisArgIleLeuValValThrAspLysGluThrIleSe 867
QY 311 -----CAGAACTCTTCTGCTCA 327
Db 867 rThrSerMetAsnHisGlnArgArgLeuValGluMetPheArgLysAsnLeuLeuGlyAr 887
QY 328 GGAATCC-----TGTTGCAAGTTCCCA-- 349
Db 887 gGluAsnAspAlaValGlnMetLysThrHisLeuArgLysLeuAlaGluGlnSerProGl 907
QY 350 -----TGCTCCCAAGCACTAGAGATGCCTCTG 378
Db 907 uGluIleGlnMetAsnLeuGlyPheSerAspPheArgProValLeuValGlnSerHisIl 927
QY 379 CTGTTCTCTTAAGAAAGATTCCAAACCAATGGTGATATGC-----CAATTGAAAGGGG 432
Db 927 eAsnGlyTyrGlnLysAspGlnLysValThrLysIleThrTyrGluGlnLeuSerAlaCy 947
```

```
QY 433 CACACAAATGCTATGATAGCAAT-----TCTAGAACAAAGAGA 471
Db 947 stleGluCysLeuIleAlaGluAsnProAlaAlaLysHisValProHisArgThrArgAl 967
QY 472 ACTA-----AAAGCACTCCATTGTTCTCAGTATCAAGATCAAAATAA 519
Db 967 aValValIleLeuArgAspLeuValLeuGlyGlnTyr----- 981
QY 520 TCTACAGTCAGATGCCCTAAACCAATGACTCTTTAGTAGGAGATTTTCCAGCATC 579
Db 982 -----ValAsnMetLeuValProThrIle 989
QY 580 AACAAAAATAATCTCATTAACAACAACCTGAGGAGCCTTACCATCGGTAGTCAACGG 639
Db 989 eThrThrTyr---ValValIleArgGlnIleHisValSerLeuAlaAlaIleLeuValTy 1008
QY 640 GTCGTCTTCCCTCGGATCAACTCTTCCAGGACCAACAATAAATACTTTGGTGGGTA 699
Db 1008 rThrGluTyrGluCysGlySerAsn-----SerSerProGlnValSerAlaSerGlyG 1026
QY 700 C-----TGAGACTGCTTTCCCACTGGGACTTTTCCACCAACTG 738
Db 1026 nValValThrCysSerThrAsnThrGlnCys---AlaSerGlyTyrThrCysAsnAsn-- 1044
QY 739 CAATTGTAATAATGTTGCAACAACCTGCATCATGATATTGAACGGTTTAAAGCCATTAA 798
Db 1045 -----GlyAlaCysCysProAsnThrAsnSerAsnThrCysSerSerAsnGlyAsnAs 1062
QY 799 GCGATGCTTGTGAATAATCCAGAGCTTCCAGCCAAAATTTGGGAAGGCCAATTGGG 858
Db 1062 nGlyCysLeu-----AlaGlyGlnThrMetVal-----AsnGlyGlnCysTy 1076
QY 859 CAATGTCAGGCCCCAGCACAAAGGGTGCACCTGAGGAGG----- 901
Db 1076 rAsn-----SerValAsnIleGlySerAlaCysGlnSerThrGlnGlnCysLeuG 1093
QY 902 -----TCAGCTGCTGAGAAATTAATCGAGTGTATGAGGCCCAAAATATGTGTCCTC 957
Db 1093 yGlySerGlnCysGlnAsnAsnIleCysGlnCysTyrSerGlyTyrValAsnValAsnG 1113
QY 958 TATTTC----- 966
Db 1113 nGlnCysValIleSerAsnGlyLeuAsnCysGlnLeuGlyThrValSerTyrAsnSerG 1133
QY 967 ATGCATT-----GGTTGCAAAAATATGAAGAAAGCCAGAACG 1005
Db 1133 nCysIleThrLeuAlaSerProGlyGlnAsnCysGlnThrSerSerGlnCysIleAspAs 1153
QY 1006 AAAGACACTAATGACATG-----CCAACTACATGCAGACTGGAGGTTT 1050
Db 1153 nSerValCysMetAsnGlnMetCysThrCysAsnAsnAsnTyrArgLeuValTyrGly-- 1172
QY 1051 GGAAGGCCCATTAACCTGCCACCAACCAATTTTCAGGACTTCCCAAGATTTCAGTCAG 1110
Db 1173 -----TyrCysValProIleThr-----SerSerIleCysGlnGlnThrGlnTh 1187
QY 1111 TAGGGGGCTTCTCATCATGCTCTCGGAGGTGTGGAGGCCACATGC----- 1159
Db 1187 rLeuValAsnAsnGlnCysValLeuSerIleValGlyGlnThrCysIleAlaAsnG 1207
QY 1160 -----CCCTGCTGCTGTCTCGAGGAGAGAGA 1185
Db 1207 nGlnCysValGlyIleAlaMetCysAsnSerGlyThrCysGlnCysThrAsnGlyAlaTh 1227
QY 1186 GCCCGAGAAGAACTGC-----TCCAAGTGCCTGGCAGACCATGAT 1230
Db 1227 rAlaMetTyrGlyTyrCysIleSerSerSerSerSerCysAsnSerAsnGlnValSe 1247
QY 1231 CCTGGAGGAATTTGGAAGGTGCTTATCACAGATT 1264
Db 1247 rIleAsn-----GlyMetCysTyrAsnThrVal 1256
RESULT 11
```

```
MMRTS
laminin beta-2 chain precursor - rat
N:Alternate names: laminin chain B3; S-laminin
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C:Accession: S03539
R:Hunter, D.D.; Shah, V.; Merlie, J.P.; Sanes, J.R.
Nature 338, 229-234, 1989
A:Title: A laminin-like adhesive protein concentrated in the synaptic cleft of the ne
A:Reference number: S03539; MUID:89159410; PMID:2922051
A:Accession: S03539
A:Molecule type: mRNA
A:Residues: 1-1801<DUM>
A:Cross-references: EMBL:X16563; NID:g57250; PIDN:CAA34561.1; PID:g57251
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
C:Function:
A:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-1801/Product: laminin beta-2 chain #status predicted <MAT>
F:36-285/Domain: VI <DOM6>
F:286-555/Domain: V <DOM5>
F:286-347/Domain: laminin-type EGF-like homology <LE01>
F:350-410/Domain: laminin-type EGF-like homology <LE02>
F:413-470/Domain: laminin-type EGF-like homology <LE03>
F:473-522/Domain: laminin-type EGF-like homology <LE04>
F:525-555/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:556-784/Domain: IV <DOM4>
F:786-831/Domain: laminin-type EGF-like homology <LE06>
F:788-1196/Domain: III <DOM3>
F:834-877/Domain: laminin-type EGF-like homology <LE07>
F:880-927/Domain: laminin-type EGF-like homology <LE08>
F:930-986/Domain: laminin-type EGF-like homology <LE09>
F:989-1038/Domain: laminin-type EGF-like homology <LE10>
F:1041-1095/Domain: laminin-type EGF-like homology <LE11>
F:1098-1143/Domain: laminin-type EGF-like homology <LE12>
F:1146-1190/Domain: laminin-type EGF-like homology <LE13>
F:1197-1412/Domain: II <DOM2>
F:1197-1412/Region: heptad repeats
F:1413-1445/Domain: alpha <ALP>
F:1446-1801/Region: heptad repeats
F:1446-1801/Domain: I <DOM1>
F:45-50/Disulfide bonds: #status predicted
F:251,371,1088,1252,1311,1351,1502/Binding site: carbohydrate (Asn) (covalent) #statu
F:1193,1196,1800/Disulfide bonds: interchain #status predicted

Alignment Scores:
Pred. NO.: 0.147 Length: 1801
Score: 118.50 Matches: 110
Percent Similarity: 27.02% Conservative: 54
Best Local Similarity: 18.12% Mismatches: 193
Query Match: 3.10% Indels: 250
DB: 1 Gaps: 29

US-09-743-237-3 (1-2134) x MMRTS (1-1801)
QY 76 CAGAGCTCAGCGCTGGAGGAGCTGCGGCTCTGTCAGGCGCCGCGCGCTGCA 135
Db 633 GlnValProGluGlnTrpAlaGluLeuValGlnArgProGlyProValSer 652
QY 136 ACGTGCACCTTCCTGCTCGCTGTACCCGCGCACCCGCGCGGCTGTTTGGCCCT 195
Db 653 AlaHisSerProCysGlyHisValLeuProArg----- 663
QY 196 GGGGGCGCTGGTCTCTGCGAGGAGGCTCCACCGCGGCTCCGATGATC----- 246
Db 664 -----AspAspArgIleGlnGlyMetLeuHisProAsnThrArgValLeuValPhePro 681
QY 247 ---CCAGTTGAATCAAGGTAAGCAGGTGGTACTACTACAAG-----TAATAA 291
Db 682 ArgProValCysLeuGluProGlyLeuSerTyrLysLeuLysLeuLysLeuThrGlyThr 701
```

```
QY 292 TCCGGAAGACAC---TTTCAGAAATCTTCTGCTCAGGAATCTCTGTG---CAAGTT 345
|||||:|||||
Db 702 GlyArgAlaHisProGluThrProTyrSerGlyLeuLeuLeuLeuLeuLeuLeu 721
QY 346 CCCATGTCGCCAGAACTAGAGATGCCCTCCCTGCTGTTCTTCTTAAGAAAGATCCCAACCC 405
|||:|||||
Db 722 ValLeuGlnProHisVal-----LeuMetLeu-----GluMetPheSerGly 735
|||||
QY 406 AATGGTGATATGCCAATTGAAAGGGGCGCACAAATGCTATGTATAGACAATCTTAGAAC 465
|||||
Db 736 GlyAspAlaAlaLeuGluArgArgThrPheGluArgTyrArg-----751
QY 466 AAGAGAACTAAAGACACTCCATTGGTTCTCCTCAGTATCAAGATCAAAATAATTATATACA 525
|||:|||||
Db 751 -----751
QY 526 GTCAGATGTCCTTAACCAATGACTGCTTTAGTAGGGAGATTTTCCAGCATCAACAAA 585
|||||:|||||
Db 752 -----CysHisGluGluGlyLeu-MetProSerLysThrPr 763
QY 586 ATTAATCTCTATTACACAACACTTCAGGAGCGCTTACCATCGGTAGTCAACGGGTCGCG 645
|||||:|||||
Db 763 OleuSer-----GluAlaCysValProLeuLeuLeuLeuLeuLeuLeuLeu 777
|||||
QY 646 TTTCCCTCGGATCAACTCTTCCAGGACCAACCAAAATAAATTGGCTGGGTACTGTGA 705
|||||
Db 777 rLeuValTyrAsnGlyAlaLeuPro-----Cysgl 787
QY 706 CTGCTTTGCCAGTGGGAGCTTTTGCACAACATGCAATTTGTAATTAATTGTTGCAACAATT 765
|||||:|||||
Db 787 nCysAspProGlnGlySerLeuSerSerGluCysAsnProHisGly-----802
QY 766 GCATCATGATATTGAACGGTTTAAAGCCATTAAAGCATGCTCTGGTAGAAATCCAGAAGC 825
|||||:|||||
Db 803 -----GlyGlnCysArgCysLysProGlyValValGlyArgArgCysAspAl 818
QY 826 TTTCCAGCAAAATTTGGAAGGGCAATTTGGCAATGTCAAGCCCCCAGCACAAACAAGG 885
|||||:|||||
Db 818 a-----CysAlaThrGlyTyrGly---PheGlyProAlaGlyCysGlnAl 833
QY 886 GTGCACTGCAGAGGTGAGGTGCTGCTGAAGAACTACTCGGAGTGTATGAGGCCCAAT 945
|||||:|||||
Db 833 aCysGlnCysSerProAspGlyAlaLeuSerAlaLeuCysGluGlyThrSerGlyGlnCys 853
|||||:|||||
QY 946 TATGCTCTCTCTCT-----958
|||||
Db 853 sLeuCysArgThrGlyAlaPheGlyLeuArgCysAspHisCysGlnArgGlyGlnTrpGl 873
QY 959 -----ATTGCAAA-----967
|||||
Db 873 yPheProAsnCysArgProCysValCysAsnGlyArgAlaAspGluCysAspAlaHisTh 893
QY 968 -----TGCATTTGGTGTGCAAAATTTATCAAGAACCCAGACGAAACACTAATGAG 1020
|||||:|||||
Db 893 rGlyAlaCysLeuGlyCysArgAspThr-----902
QY 1021 CATGCCAACTACATGCAGCTGGAGGT-----TTGGAAGGAGCA 1062
|||||:|||||
Db 903 -----ThrGlyGlyGluHisCysGluArgCysIleAlaGlyPheHi 916
QY 1063 TTACTTGCACCAACAATAATTTTCAGGACTTCCAAGATTCAGTCAGATAGCGGCTTC 1122
|||||:|||||
Db 916 sGlyAspProArgLeuProTyrGlyGlyGlnCysArg-----928
QY 1123 CTCATGCATCTCTCGGAGGTGGTGGAGCCACATCGCCTGCTTGTCTCAGGGAGA 1182
|||||:|||||
Db 929 -----ProCysProCys-----ProGluGlyPr 936
QY 1183 AGAGCGCGAGAAAGAACTGCTCCAAGTGCCTGGCAGCAGAGATGATCTCGGAGGAAT 1242
|||||:|||||
Db 936 oGlySerGlnArgHisPheAlaThrSerCysHisArgAsp-----949
QY 1243 TGAAGGTGCTTATCACAGATTCTCCACACTGAGTTTAAATCTAAGGGATTGAAATGGA 1302
```

```
Db 950 -GlyTyrSerGlnGlnIleValCysHisCysArgAlaGlyTyrThrGlyLeuArgCysgl 969
|||||:|||||
QY 1303 GTAGAGTATAAG-----TG 1317
Db 969 uLaCysAlaProGlyHisPheGlyAspProSerLysProGlyGlyArgCysGlnLeuCy 989
QY 1318 TGAATGCATGTTGTAUTTTTCTTCTAGTCTAGAAATCTCTAGTTTAGAAAGATGTTTTAGGG 1377
|||||
Db 989 sGluCysSer-----Glu 993
QY 1378 GAACATGAGGCTGGCTCTGCAGCAACACACAGGCTCCCTCGCATCCCTGGGCCA-----1432
|||||:|||||
Db 993 yAsnIle-----AspProThrAspProGlyAlaCysAspProHisTh 1007
QY 1433 -GGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCATGCCCTTTTCTGAGGAGG 1491
|||||:|||||
Db 1007 rGlyGlnCysLeuArgCysLeuHisHisThrGluGlyPro-----1020
QY 1492 TGCATGGCCTGAGCATTTGTTGCTGGCCAGAGGAGAGAGCTTGGTTCCCATAGTCCT 1551
|||||
Db 1021 -----HisCys-----1022
QY 1552 GGGAGAGTCTCTGCAGGCGCGGAGGCGAGAGC-----AGCCCTCGGAGAGCTCAC 1605
|||||:|||||
Db 1023 -GlyHisCysLysProGlyPheHisGlyGlnAlaAlaArgGlnSerCysHisArgCysTh 1042
QY 1606 T-----CTGGTCGACTCTCTCTCAGAGAATGTGTCTGAGAGCTCTCTGCAATGAA 1659
|||||:|||||
Db 1042 rCysAsnLeuGlyThrAspProGlnArg-----CysProse 1055
QY 1660 AACCTTAATGGTTCTTCTGT 1678
Db 1055 rThrAspLeuCysHisCys 1061
```

RESULT 12

T42215

zonadhesin - mouse

N:Alternate names: sperm-specific membrane protein

C:Species: Mus musculus (house mouse)

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T42215

R:Gao, Z.; Garbers, D.L.

J. Biol. Chem. 273, 3415-3421, 1998

A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane

A:Reference number: 222080; MUID:98123114; PMID:9452463

A:Accession: T42215

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-5376 <GAO>

A:Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1

C:Genetics:

A:Gene: Zan

A:Map position: 5

C:Function:

A:Description: functions in multiple cell adhesion processes

A:Note: found exclusively on the apical region of the sperm head

C:Keywords: cell adhesion

Alignment Scores:

Pred. NO.: 0.151

Score: 118.50

Percent Similarity: 29.39%

Best Local Similarity: 20.03%

Query Match: 3.10%

DB: 2

US-09-743-237-3 (1-2134) x T42215 (1-5376)

Length: 5376

Matches: 137

Conservative: 64

Mismatches: 232

Indels: 251

Gaps: 41

QY 16 CGAAGCTCCGGGGGGCGGCGAGCAGCGCGGAGCTCTCGGGGAGTACCCGGGATCC 75

Db 3478 LysAspAlaGlnGlyAlaLeuLeuProAlaGlyLysThrTrpThrSerProGlyCysThr 3497

A:Accession: A54148
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2406 <LEV>
C:Genetics:
A:Gene: FlyBase:Ten-m
A:Cross-references: FlyBase:FBgn000449
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:514-540/Domain: EGF homology <EGF>
F:610-637/Domain: EGF homology <EGF1>

Alignment Scores:
Pred. No.: 0.163 Length: 2406
Score: 118.00 Matches: 132
Percent Similarity: 31.7% Conservative: 79
Best Local Similarity: 19.8% Mismatches: 240
Query Match: 3.09% Indels: 214
DB: 2 Gaps: 35

US-09-743-237-3 (1-2134) x A54148 (1-2406)

```
QY 7 GGTCAAGCGGAGCTCGCGGGGGGACACA---GCCAGCGGGGGAGCTCCTCGGGGAGT 63
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 GlySerGlyGlySerSerAlaAlaThrValThrAlaThrSerAsnSerGlyThr 335
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 ACCCGGGATCCAGAGCTCAGCGCGCTGGAGGAGCTCGCGCTCTGCAGGCCCGCCAGC 123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 336 AlaGlnGlyLeuGlnSer-----ThrSerAlaSerAlaGluAlaThrSer 350
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 CGCCCGCTGCGAACGTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 351 -----SerAlaAlaThrSer-----SerSerGlnSer 359
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 TGTTCCTGCGCGCGCTCGGTCTCTGCGAAGAGCGCTCCACCGCGGGCTCCCGATG 243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 SerLeuThrProSerLeuSerSerSerLeuAlaAlaAsnAsnGlyGlyAlaArgThr 379
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 ATCCAGTTGAATCAAGTAAAGCAGCTGTCTACTACTACTACTACTACTACTACTACTACT 303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 PheProAla-ArgSerPheProAspGlyThrPheGly-----GlnI 395
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 ACTTTGCGAATCTTCTGCTCAGAAATCTGTTCCAAATGTCCTCCAGTCCCGAGAACT 363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 395 eThrLeuGlyGlnLysLeuThrLysGlnIleGlnProTyrSerTyrTrpAsnMetGlnPh 415
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 AGAGGATGCTCTGCTGTTCTCTTAAGAAAGAT----- 397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 eTyrGlnSerGluProAlaTyrValLysPheAspTyrThrIleProArgGlyAlaSerI 435
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 398 -----TCCAAACCAATGGTGATATGCCAATTGAAAGGGGCGACACAAAT 441
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 eGlyValTyrGlyArgArgAsnAlaLeuProThrHisThrGlnTyrHisPheLysGluVa 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 442 GCTATGTATAGACAATCTTAGACAGAGAACTAAAGCACTCCATTTG----- 490
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 lLeuSerGlyPheSerAlaSerThrArgThrAlaArgAlaAlaHisLeuSerIleThrAr 475
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 491 ----GTTCTCTAGTACAAGTCAAAATAATATTACTACAGTCA----- 529
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 gGluValThrArgTyrMetGluProGlyHisTrpPheValSerLeuTyrAsnAspGlu 495
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 -GATGTCCTAAACCAATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAACAAAT 588
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 495 yAspValGlnGlu-----LeuThrPheTy 503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 589 AAATCTATTACACAACTTAGGGAGGCTTACCATCGGTAGTCAAGGGTCTGCTTT 648
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 503 rAlaAlaValAlaGluAspMetThrGlnAsnCysProAsnGlyCysSerGlyAsnGly-- 522
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 649 CCCTCGGGATCAACTCTTCCAGGACCACCAAAATAAATTTGGCTGGCTACTGTGCTAC 708
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 -----GlnCysLeuLeuGlyHisCysGlnCy 531
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 709 CTTTGGC-----AGTGGGAGCTTTTGCACAACTGCAATTGTAATAATTGTTGCAACA 762
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 531 sasnProGlyPheGlyAspAspCysSerGluSerValCysProValLeuLeuCysSerGl 551
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 763 CTTGCATCATGATATTAACGGTTTAAAGCCATTAAAGCATTTCTGGTAGAATCCAGA 822
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 n---HisGlyGluTyrThrAsnGlyGlyCysIle-----Cys-----AsnProGl 565
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 823 AGCTTTCCAGCCAAATAATTGGGAGGGCCAAATTGGCAATGTCAAGCCCCCAGCACAA 882
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 y-----TriLysGlyLysGluCysSerLeuArgHisAspGluCysGl 579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 883 AGGGTGCAACTGCAGGAGGTGAGC---TGCCTGAAGAATACTACTGCGAGTGC--- 931
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 uValAlaAspCysSerGlyHisGlyHisCysValSerGlyLysCysGlnCysMetArgGl 599
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 -TATGAGGCCCAAAATTATGTCTTCTATT---TGC-----AAATGCAATGCT-- 976
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 yTyrLysGlyLys---PheCysGluGluValAspCysProHisProAsnCysSerGlyH 618
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 977 -----TCCAAAATTTATGAAGAAGCCCGACAGCAAA 1008
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 618 sGlyPheCysAlaAspGlyThrCysIleCysLysLysGlyTrpLysGlyProAspCysAl 638
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1009 GACA-----CTAATGAGCATGCCAAACTACATGCACAGCTGGAGTTTGA 1053
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 638 aThrMetAspGlnAspAlaLeuGlnCysLeuProAsp----- 650
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1054 AGCAGCCATTACCTGCCACCAAGAAATTTTCAGGACTTCCAAAGATTTCAGTACAGTAG 1113
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 651 -----CysSerGlyHisGlyThrPheAspLeuAspTh 661
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1114 GCGGCTTCTCATGCATCTCTGGGAGGTGGTGGAGGCCACATCGCCTGCTGCTTGC 1173
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 rGln-----ThrCysThrCysGluAlaL 669
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1174 T---CAGGGAGAGAGCGCGAGAAAGAACACATGC-----TCCAA 1209
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 669 sTrpSerGlyAspAspCysSerLysGluLeuCysAspLeuAspCysGlyGlnHisGlyAr 689
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1210 GTGCTGGCAGACAGATGATCTGGAG---GAATTTTGA-----AGTGCTTATCACA 1260
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 689 gCysGluGlyAspAlaCysAlaCysAspProGluTrpGlyGlyGlyCysAsnThrAr 709
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1261 GATTCCTCCACACTGACTTTAAATCTAAGGATTGAAATGGAGTACAGATATAAGTGA 1320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 709 gLeuCysAspValArgCysAsnGluHisGly-----GlnCysL 722
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1321 A-----TGCATGTTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAGGATGT 1371
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 722 sAsnGlyThrCysLeuCysValThrGlyTrpAsn----- 733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1372 TTAGGGACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCTGCTCATCCTGGGCC 1431
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 734 -----GlyLysHisCysThrIleGluGlyCysProAsnSerCysAlaGlyHisGly-- 750
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1432 AGGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTCTTCTGAGGAG 1491
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 -----GlnCysArgValSerGlyGluGl 758
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1492 TCATGCGCTGACCATTTGTTCTTGGCCCGAGAGAGAGCTTTGGTTTCCCATGCTCCT 1551
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 758 yGlnTrpGluCysArgCysTyGluGly-----TipAspGlyProAspCy 773
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1552 GGGAGTCTTCGAGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1611
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 773 sGlyIleAlaLeuGluLeuAsnCysGlyAspSerLysAspAsnAspLysAspGlyLeuVa 793
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1612 CGACTCTTCTCTCAGAGAAATGTTGCTCTGGAGGCTGCTGCTGCAAAACCCCTAATGCT 1671
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 793 lAspCysGluAspPro-GluCysCysAlaSerHis----- 804
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1672 TTCTTGTGTTTTCAAAATTTATTTAGAAATAAGTTCTCCGGATGGGCTGTTGTGATACC 1731
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 805 -----ValCysLysThrSerGlnLeuCysValSerAlaPro----- 816
 Qy 1732 ACTTAAATCTCTAGAGACTACTGAACACCCCTAAAGATTTTCTCTAGCGTAGATATTTC 1791
 Db 817 -----LysProIleAspValLeuLeuArgLysGlnProp 828
 Qy 1792 CCAGAGACAGC 1802
 Db 828 roAlaIleThr 831
 RESULT 14
 S47008
 tenascin-like protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C>Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S47008
 R:Baumgartner, S.; Martin, D.; Hagios, C.; Chiquet-Ehrismann, R.
 EMBO J. 13, 3728-3740, 1994
 A:Title: Ten(m), a *Drosophila* gene related to tenascin, is a new pair-rule gene.
 A:Reference number: S47008; MUID:94349920; PMID:8070401
 A:Accession: S47008
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2515 <BAU>
 A:Cross-references: EMBL:X73154; NID:g510505; PIDN:CAA51678.1; PID:g510506
 C:Genetics:
 A:Gene: FlyBase:Ten-m
 A:Cross-references: FlyBase:Fgn0004449
 C:Superfamily: unassigned EGF-related proteins; EGF homology
 F:298-324/Domain: EGF homology <EGF>
 F:394-421/Domain: EGF homology <EGF1>
 Alignment Scores:
 Pred. No.: 0.163 Length: 2515
 Score: 116.00 Matches: 132
 Percent Similarity: 31.73% Conservative: 79
 Best Local Similarity: 19.85% Mismatches: 240
 Query Match: 3.09% Indels: 214
 DB: 2 Gaps: 35
 US-09-743-237-3 (1-2134) x S47008 (1-2515)
 Qy 7 GGTCAAGCGAAGCTCGCGGGGGCGACA---CGCAGCGCGGGAGCTCTCGGGAGT 63
 Db 100 GlySerGlyGlySerSerAlaAlaThrValThrAlaThrSerAsnSerGlyThr 119
 Qy 64 ACCCGGGATCCAGAGCTCAGCGCGTGGAGAGCTCGCGCTCTCGAGGCCCGCGAGC 123
 Db 120 AlaGlnGlyLeuGlnSer-----ThrSerAlaSerAlaGluAlaThrSer 134
 Qy 124 CGCCGCGCTGCAAGCTGCACCTTCCTGCTCCTGCTACCGCGCAGCGCGCGGG 183
 Db 135 -----SerAlaAlaThrSer-----SerSerGlnSer 143
 Qy 184 TGTTTTCGCCCTGGCGCCCTGGTCTCGCAGAGAGCCCTCCACCGCGCGCTCCGCATG 243
 Db 144 SerLeuThrProSerLeuSerSerSerLeuAlaAlaAsnAsnGlyGlyAlaArgThr 163
 Qy 244 ATCCAGTTGNAATCAGGTAAAGCAGGTGGTACTACTACAGTAAATATCCGGNAGAGC 303
 Db 164 PheProAla-ArgSerPheProAspGlyThrThrPheGly-----GlnIle 179
 Qy 304 AACTTTGCAGAAATCTTCTGCTCAGGAATCCTGTTGCAAGTTCCTCAGTCCAGGAACT 363
 Db 179 eThrLeuGlnLysLeuThrLysGluIleGlnProTyrSerTyrTrpAsnMetGlnPh 199
 Qy 364 AGAGATGCTCCGCTGTTCTCTTAAGAAGAT----- 397
 Db 199 eTyrGlnSerGluProAlaTyrValLysPheAspTyrThrIleProArgGlyAlaSerIle 219
 Qy 398 -----TCCACCAATGTTGATATGCAATTTGAAGGGGCGCACAAAT 441
 Db 506 sasnglythrCysLeuCysValThrGlyTrpAsn----- 517

Db 219 eGlyValTyrGlyArgAsnAlaLeuProThrHisThrGlnTyrHisPheLysGluVa 239
 Qy 442 GCTATGATAGACAATTTCTAGACAAGAACTAAAGACACTCCATTTG----- 490
 Db 239 lLeuSerGlyPheSerAlaSerThrArgThrAlaArgAlaAlaHisLeuSerIleThrAr 259
 Qy 491 -----GTTCTCAGTATCAAGATCAAAATAATTATCTACAGTCA----- 529
 Db 259 gGluValThrArgTyrMetGluProGlyHisTrpPheValSerLeuTyrAsnAspSpGl 279
 Qy 530 -GATGCTCCCTAAACCAATGACTGCTTTAGTAGGAGATTTTGGCCAGCATCAACAAAT 588
 Db 279 yAspValGlnGlu-----LeuThrPheTy 287
 Qy 589 AAATCTCATTACACAACAACCTTACCGAGAGCTTACCATCGGTAGTCAACGGGTCTGCTTT 648
 Db 287 rAlaAlaValAlaGluAspMetThrGlnAsnCysProAsnGlyCysSerGlyAsnGly-- 306
 Qy 649 CCCTCGGGATCAACTCTTCAGGACCACCACAAATAAATTAATTTGGCTGGTACTGTGACTG 708
 Db 307 -----GlnCysLeuLeuGlyHisCysGlnCy 315
 Qy 709 CTTTGGC-----AGTGGGACCTTTGCAACAACCTGCAATTTGTAATAATTTGTTGCAACA 762
 Db 315 sasnProGlyPheGlyGlyAspCysPysSerGluSerValCysProValLeuCysSerGl 335
 Qy 763 CTTGCATCATGATGATGAACGGTTTAAAGCCATTAAGGCATGCTTGGTGAAGATCCAGA 822
 Db 335 n---HisGlyGluTyrThrAsnGlyGluCysIle-----Cys-----AsnProGl 349
 Qy 823 AGCTTTCCAGCCCAAAATTTGGGAAGGCCAAATTTGGGAATGTCAAGCCCCAGCACAAACA 882
 Db 349 y-----TrpLysGlyLysGluCysSerLeuArgHisAspGluCysGl 363
 Qy 883 AGGTCGCACTGAGGAGGTGAGC---TGCTGCAAGAANTTACTGCGAGTGC----- 931
 Db 363 uValAlaAspCysSerGlyHisGlyHisCysValSerGlyLysCysGlnCysMetArgGl 383
 Qy 932 -TATGAGGCCCAATTTATGTTCTTCTTATT---TGC-----AAATGCTATGCT-- 976
 Db 383 yTyrLysGlyLys---PheCysGluGluValAspCysProHisProAsnCysSerGlyHi 402
 Qy 977 -----TGCAAAATTTATGAAGAACCCAGACAGCAAA 1008
 Db 402 sGlyPheCysAlaAspGlyThrCysIleCysLysLysGlyTrpLysGlyProAspCysAl 422
 Qy 1009 GACA-----CTAATGAGCATGCCAAACTACATCGACAGTGGAGGTTTGA 1053
 Db 422 aThrMetAspGlnAspAlaLeuGlnCysLeuProAsp----- 434
 Qy 1054 AGCAGCGCATTTACCTGCCACCAACGAAATTTTCAGGACTTCCAGATTCACGATAG 1113
 Db 435 -----CysSerGlyHisGlyThrPheAspLeuAspTh 445
 Qy 1114 GCGGCTTCTCATGCATCTCCTGGGAGGTGGTGGAGCCACATGCGCTGCTGCTGTC 1173
 Db 445 rGln-----ThrCysThrCysGluAlaLy 453
 Qy 1174 T---CAGGGAGAACAGGCCCGAGAAAGAACACTGC-----TCCAA 1209
 Db 453 sTrpSerGlyAspAspCysSerLysGluLeuCysAspLeuAspCysGlyGlnHisGlyAr 473
 Qy 1210 GTGCTGGCAGAGCAGATGATCTCTGGAG---GAATTTGGA-----AGGTGCTTATCACA 1260
 Db 473 gCysGluGlyAspAlaCysAlaCysAspProGluTrpGlyGlyClyTrpCysAsnThrAr 493
 Qy 1261 GATTCTCCACACAGTGTAAATCTAAGGATTTGAATCGAGTAGAGTATAGAAGTGA 1320
 Db 493 gLeuCysAspValArgCysAsnGluHisGly-----GlnCysLy 506
 Qy 1321 A-----TGCATGTTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAGGATGT 1371
 Db 506 sasnglythrCysLeuCysValThrGlyTrpAsn----- 517

QY 1372 TTAGGGAAACATGAGGCTGGCTCTGCAGCAACAAACAGGCTCCCTCGCATCCCTGGGCC 1431
 Db 518 -----GlyLysHisCysThrIleGluGlyCysProAsnSerCysAlaGlyHisGly-- 534
 QY 1432 AGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTTCTGAGGAGG 1491
 Db 535 -----GlnCysArgValSerGlyGluG1 542
 QY 1492 TGCATGGCCTGAGCATTGTTGCTCGGCCAGAGGAGAGCTTGGTTCCTCCATAGTCCT 1551
 Db 542 yGlnTrpGluCysArgCysTyrGluGly-----TrpAspGlyProAspCys 557
 QY 1552 GGGAGAGTGTCTGCAGCGGGCGGAGGAGCAGAGCAGGCTCGGGAGAGCTCACTCTGTT 1611
 Db 557 sGlyIleAlaLeuGluLeuAsnCysGlyAspSerLysAspAsnAspLysAspGlyLeuVa 577
 QY 1612 CGACTCTTCTCAGAGAAATGTTCTGAGGAGCTGCTGCTGCATGAAACCCCTAATGTT 1671
 Db 577 lAspCysGluAspPro-GluCysCysAlaSerHis----- 588
 QY 1672 TTCTCTGTTGTTTTCAAATTATTAGAAATAGTTCTCGGATGGCTGTTGTGATACC 1731
 Db 589 -----ValCysLysThrSerGlnLeuCysValSerAlaPro----- 600
 QY 1732 ACTTAAATCTCTAGAGAACTACTGAACACACTTCTAGCGTAGATATTTC 1791
 Db 601 -----LysProIleaspValLeuLeuArgLysGlnProp 612
 QY 1792 CCAGAGACACG 1802
 Db 612 roAlaIleThr 615

RESULT 15

T34513

Hypothetical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34513

R:Pavello, A.; Vaudin, M.

Submitted to the EMBL Data Library, August 1994

A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: Z21536

A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3507 <FAV>

A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CESP:ZK783.1

A:Map position: 3

A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
 3504/1

Alignment Scores:

Pred. No.:	0.181	Length:	3507
Score:	117.50	Matches:	117
Percent Similarity:	28.57%	Conservative:	67
Best Local Similarity:	18.17%	Mismatches:	170
Query Match:	3.07%	Indels:	290
DB:	2	Gaps:	33

US-09-743-237-3 (1-2134) x T34513 (1-3507)

QY 8 GGTCAAGCGAAGCTCGCGGGGGGCGACAGCGCGGGAGCTCCTCGGGAGTACCC 67
 Db 1268 GlySerGlyGluAla-----GlyValPro 1275
 QY 68 CGGGATCCAGAGCTCAGCGCGCTGGAGGAGCTGGCGCTCCCTGCAGGC-----CCCGCA 121
 Db 1276 GluSerProAspGlySerGlyGluSerThrSerAlaProAspGlyValSerProThr 1295

QY 122 GCGCCCGCTGCAACGTCACACTTCCTGCTGCTGCTACCCGCCACCG----- 172
 Db 1296 SerSerAlaThrAlaProGluValProThrThrSerAlaSerThrProAspAlaVal 1315
 QY 173 CAGCCCGCGGGTGT-----TTGCCCTGGCGGCCCTGGGT 208
 Db 1316 GluGluSerGlyIleProSerThrSerLysProThrAlaGluProLeuGluThrAla 1335
 QY 209 CTGCGCAAGAG-----CCTCCC 226
 Db 1336 ProSerThrGluValThrSerProGluGlySerGlyThrGluGluSerThrLeuProPro 1355
 QY 227 ACCCGGCGCTCGCATGATCCAGTTGAAATCAAGGTAAGCAAGAGTGGTACTACTACAAGT 286
 Db 1356 ThrGluGlySer-----GlyGluSerThrThrSer 1365
 QY 287 ATAATCCG-----GAAGAAGCAACTTTCAGAAATCTTCTGCTCAGGAATCCTGTGC 340
 Db 1366 SerAlaProThrValGluProAlaThrVal----- 1375
 QY 341 AAGTTCCCATGGTCCCGAGGAAGTAGAGGATGCCTGCTGCTTCTTAAGAAAGATTCC 400
 Db 1375 ----- 1375
 QY 401 AACCCAAATGGTATATGCCAAATTGAAAGGGGCACACAAATGCTATATAGACAATCT 460
 Db 1375 ----- 1375
 QY 461 AGAACAAGAGAACTAAAGACACTCCATTGGTTTCTCAGTATCAAGATCAAAATAATTAT 520
 Db 1376 -----LeuProGlnAsnArgAsnGluLys----- 1384
 QY 521 CTACAGTCAGATGCTCCCTAAACAATGACTGCTTTAGTAGGGAGATTTTTCAGCAGCATCA 580
 Db 1385 -----ProGluPro 1387
 QY 581 ACAAAATAAATCTATTACACAACACTTGAGGGAGCCTTACCATCGGTAGTCAACGGG 640
 Db 1388 ThrLysAspThrPhe-----AlaLeuProThrThrThrGly 1400
 QY 641 TCTGCTTTCCCTCGGGATCAACTCT----- 667
 Db 1401 AlaProGlnAlaAsnAspSerSerValGluAsnThrLysCysThrSerSerAspGluCys 1420
 QY 668 -----CCAGCA----- 673
 Db 1421 GlyLeuAspAlaLeuCysGluArgArgThrGlyValCysArgCysGluProGlyPheGlu 1440
 QY 674 -----CCACCAAAATAAATTTGGCTGGTACTGTGACTGCTTTCAGTGGGACTTTT 727
 Db 1441 GlyAlaProProLysLysSerCysValAspValAspGluCys---AlaThrGlyAsp--- 1458
 QY 728 TGCACAACACTGCAATTTGTAATAATTGTCACAACACTTGCATCATGATATTGAACGGTTT 787
 Db 1459 ---HisAsnCysHisGluSerAlaArgCysGlnAsn-----Tyr 1470
 QY 788 AAAGCCATTAAAGCATGCTCTTGTAGAAATCCAGAAAGCTTTCAGCCAAAATTTGGGAAG 847
 Db 1471 ValGlyGlyTyrAlaCysPhe-----CysProThrGlyPheArg---LysAlaAspAsp 1487
 QY 848 GGCCAATTGGCAATGTCAAG-----CCCCAGCACAAAGGGTGCAACTGCAGGAGG 901
 Db 1488 GlySerCysGlnAspIleaspGluCysThrGluHisAsnSerThrCys----- 1503
 QY 902 TCAGGCTGCTGAAGAATTACTCGAGTGCTATGAGGCCCAAAATTATGTTCTTCTATT 961
 Db 1504 -----CysGlyAlaAsn 1507
 QY 962 TGCAAATGCTTGGT-----TGCAAAATTTATGAAGAAGC 997
 Db 1508 AlaLysCysValAsnLysProGlyThrThrSerCysGluCysGluAsn----- 1523
 QY 998 CCAGAACGAAAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGGTTTGAAGGC 1057

```
Db 1524 -----GlyPheLeuGly 1527
Qy 1058 AGCCATTAC-----CTGCCACCAAGAAATTTTCAGGACTTCCAAGATTTCAGTCACGAT 1111
Db 1528 AspGlyTyrGlnCysValProThrThrLys-----LysProCysAspSerThrGln 1544
Qy 1112 AGCGCGCCTTCCTCATGCATCTCCTGGAGGTGTGGAGGCACATGCGCCTGCTGCTT 1171
Db 1545 SerSerLysSerHisCys-----SerGluSerAsnMetSerCysGluVal 1559
Qy 1172 GCTCAGGAGAGAGAGCGGAGAGACACTGCTCCAAGTCCCTGGCA-----1219
Db 1560 AspThrValAspGlySerValGlu-----CysLysGluCysMetGlyGlyTyrLysLys 1577
Qy 1220 GAGCAGATGATCCTCGAGGATTTGGAAGTGTCTATACAGATTCTCCACACTGAGTTT 1279
Db 1578 SerGlyLysValCysGluAspIleAsnGluCysValAlaGluLysAlaProCysSerLeu 1597
Qy 1280 AAATCTAAGGGATTGAAATGGAGTAGAGTATAAGTGTGAATGCATGTTGATTTGTCT 1339
Db 1598 AsnAlaAsnCysValAsnMetAsnGlyThrPheSerCysSerCys-----1612
Qy 1340 TAGTCTAGAAATCTCTAGTTTAGAAGGATGTTTAGGGGAACATGAGCTGGCTGTGCAG 1399
Db 1613 -----LysGlnGly-TyrArgGlyAsp---GlyPheMetCysTh 1624
Qy 1400 C-----AACACCAGGCTCCCTGCTGCATCCCTGGGCCAGGAGTTTACTC 1444
Db 1624 rasPIleAsnGluCysAspGluArgHisProCysHisProHisAlaGluCysThrAsnLe 1644
Qy 1445 AGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTCT-----1484
Db 1644 uGluGlySerPheLysCysGluCysHisSerGlyPheGluGlyIleLysLysCy 1664
Qy 1485 -----GAGGAGTGCATGGCCTG-----AG 1504
Db 1664 sThrAsnProLeuGluArgSerCysGluAspValGluLysPheCysGlyArgValAspHi 1684
Qy 1505 CATTTGTTTCTTGGCCAGAGAGAGAGCTTGGGTTCCTAGTCCCTGGAGAGTGTCTG 1564
Db 1684 sValSerCysLeuSerValArgIleTyrAsnGly-----SerLeuSerSerValCy 1701
Qy 1565 CAGGCGCGGAGGAGAGAGAGGCGCTGCGGAGAGTCTACTCTGGTCGACTTCTCTCT 1624
Db 1701 sGluCysGluProGlyPheArgPheGluLysGluSerAsn-----SerCysValAspIl 1719
Qy 1625 CAGAGATGT 1634
Db 1719 eAspGluCys 1722
```

Search completed: April 21, 2003, 12:03:50
Job time : 118.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:50:04 ; Search time 30.5 Seconds
(without alignments)
5803.966 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 3824
Sequence: 1 aattcgggtcaagcgcaag.....aaaaaaaaaaaaaaaaaa 2134

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=rlp
-Q/cgn2_1/USPTO.spool/US09743237/runat_21042003_112209_1461/app_query.fasta_1.2311
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09743237 -CNCN_1_1_47 -runat_21042003_112209_1461 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1647	43.1	299	1 MTL5_HUMAN	Q9v415 homo sapien
2	1251	32.7	295	1 MTL5_MOUSE	Q9wtj6 mus musculu
3	130.5	3.4	3695	1 LMA5_HUMAN	Q15230 homo sapien
4	123	3.2	1700	1 BAR3_CHITE	Q03376 chironomus
5	119.5	3.1	1172	1 AHM2_ARATH	O64474 arabidopsis
6	119.5	3.1	3010	1 POLG_HCVJA	P26662 h genome po
7	119	3.1	1576	1 YLK3_CAEEL	P41951 caenorhabdi
8	118.5	3.1	1801	1 LMB2_RAT	P15800 rattus norv
9	118.5	3.1	5376	1 ZAN_MOUSE	O88799 mus musculu
10	117	3.1	3011	1 POLG_HCV1	P26664 h genome po
11	117	3.1	3718	1 LMA5_MOUSE	O61001 mus musculu
12	116.5	3.0	2003	1 NTC4_HUMAN	Q99466 homo sapien
13	116	3.0	3010	1 POLG_HCVBK	P26663 h genome po
14	116	3.0	3010	1 POLG_HCVJT	Q00269 h genome po
15	115.5	3.0	769	1 ITB2_HUMAN	P05107 homo sapien
16	114.5	3.0	778	1 TAST_HUMAN	Q12815 homo sapien
17	114.5	3.0	3726	1 TRX_DROME	P20659 drosophila
18	112.5	2.9	3133	1 HMCT_BOMBO	P98092 bombyx mori

ALIGNMENTS

RESULT 1

ID	MTL5_HUMAN	STANDARD;	PRT;	299 AA.
AC	Q9v415;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).			
DE	MTL5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=99208669; PubMed=10191092;			
RA	Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;			
RT	"A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation."			
RL	Genomics 57:130-136(1999).			
CC	-!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.			
CC	-!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; U86074; RAD24668.1;			
DR	Genew; HGNC:7446; MTL5.			
DR	MTM; 604374;			
DR	InterPro; IPR005172; CXC.			
DR	Pfam; PF03638; CXC; 1.			
DR	Spermatogenesis.			
KW	DOMAIN 99			CYS-RICH.
FT	SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;			

Alignment Scores:

```
Pred. No.: 9,84e-121 Length: 299
Score: 1647.00 Matches: 299
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.07% Indels: 0
DB: 1 Gaps: 0

US-09-743-237-3 (1-2134) x MTL5_HUMAN (1-299)
QY 407 ATGGTGATATGCCAATTCGAAGGGGGACACAAATGCTATGATAGACAATCTAGACA 466
Db 1 MetValIleCysGlnLeuLysGlyGlyThrGlnMetLeuCysIleAspAsnSerArgThr 20
QY 467 AGAGAACTAAAGCACTCCATTTGGTCTCAGTATCAAGATCAAAATAATATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuValProGlnTyrGlnAspGlnAsnTyrLeuGln 40
QY 527 TCAGATGTCCTTAACCAATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAACAAA 586
Db 41 SerAspValProLysProMetThrAlaLeuValGlyArgPheLeuProAlaSerThrLys 60
QY 587 TTAATCTCATTAACAACAACCTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnGlnLeuGlyAlaLeuProSerValValAsnGlySerAla 80
QY 647 TTCCTCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGAC 706
Db 81 PheProSerGlySerThrLeuProGlyProLysIleThrLeuAlaGlyTyrCysAsp 100
QY 707 TCGTTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTAATTTTGGCAACACTTG 766
Db 101 CysPheAlaSerGlyAspPheCysAsnAsnCysAsnCysAsnAsnCysCysAsnLeu 120
QY 767 CATCATGATATTGAACGGTTTAAAGCCATTAAAGGCATCTTGGTGGAGAAATCCAGAACT 826
Db 121 HisHisAspIleGluArgPheLysAlaLeuLysAlaCysLeuGlyArgAsnProGluAla 140
QY 827 TTCACGCCAAAATAATGGGAAGGCCAATTTGGGCAATGTCAGCCCAAGCCACACAAAGG 886
Db 141 PheGlnProLysIleGlyLysGlyGlnLeuGlyAsnValLysProGlnHisAsnLysGly 160
QY 887 TCGCACTCGAGAGTCAAGCTGCTGAAGATTAATCTCGAGTGTATGAGCCCAATTT 946
Db 161 CysAsnCysArgSerGlyCysLeuLysAsnTyrCysGlyCysTyrGluAlaGlnIle 180
QY 947 ATGCTCTCTCTATTGCAATGCAATGCTTGGTCAAAATATTGAAGAAAGCCCAAGCA 1006
Db 181 MetCysSerSerIleCysLysCysIleGlyCysLysAsnTyrGluGluSerProGluArg 200
QY 1007 AAGACACTAATGAGATGCCAACTACATGCAAGTGGAGTGGTGGAGGAGGAGCAATTTAC 1066
Db 201 LysThrLeuMetSerMetProAsnTyrMetGlnThrGlyGlyLeuGluGlySerHisTyr 220
QY 1067 CTGCCACCAACGAATTTTCAGGACTTCCAGATTCAGTACGATAGGCGGCTTCCTCA 1126
Db 221 LeuProProThrLysPheSerGlyLeuProArgPheSerHisAspArgArgProSerSer 240
QY 1127 TGCATCTCTGGAGGTGGTGGAGGCCACATGCGCTCGCTTGTCTGCTCAGGAGAGAG 1186
Db 241 CysIleSerTrpGluValValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 260
QY 1187 GCCGAAAGAACACTGCTCCAAGTGGCTGGCAGACAGATGATCTCTGGAGGAATTTGGA 1246
Db 261 AlaGluLysGluHisCysSerLysCysLeuAlaGluGlnMetIleLeuGluGluPheGly 280
QY 1247 AGGTCTTATCACGATTTCTCCACTGAGTTAAATCTAAGGAGTTGAAATGAG 1303
Db 281 ArgCysLeuSerGlnIleLeuHisThrGluPheLysSerLysGlyLeuLysMetGlu 299

RESULT 2
MTL5_MOUSE
ID MTL5_MOUSE STANDARD: PRT: 295 AA.
```

```
Q9WTJ6;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
metalothionein-like protein).
MTL5
Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U77383; AAD24667.1; -
CC EMBL; U67176; AAD24666.1; -
CC MGD; MGI:1340029; MTL5
CC InterPro; IPR005172; CXC.
CC Pfam; PF03638; CXC; 1.
CC Spermatogenesis.
CC DOMAIN 99 187 CYS-RICH.
CC SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Alignment Scores:
Pred. No.: 7,46e-90 Length: 295
Score: 1251.00 Matches: 228
Percent Similarity: 84.95% Conservative: 26
Best Local Similarity: 76.25% Mismatches: 41
Query Match: 32.71% Indels: 4
DB: 1 Gaps: 1

US-09-743-237-3 (1-2134) x MTL5_MOUSE (1-295)
QY 407 ATGGTGATATGCCAATTCGAAGGGGGACACAAATGCTATGATAGACAATCTAGACA 466
Db 1 MetValIleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysIleAspAsnCysGlyAla 20
QY 467 AGAGAACTAAAGCACTCCATTTGGTCTCCTCAGTATCAAGATCAAAATAATATCTACAG 526
Db 21 ArgGluLeuLysAlaLeuHisLeuProGlnTyrAspGlnSerPheProGln 40
QY 527 TCAGATGTCCTTAACCAATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAACAAA 586
Db 41 SerGluLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLys 60
QY 587 TTAATCTCATTAACAACAACCTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCT 646
Db 61 LeuAsnLeuIleThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAla 80
QY 647 TTCCTCGGGATCAACTCTTCCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGAC 706
Db 81 PheProSerGlyProAlaLeuGlnGlyProLysIleThrLeuSerGlyTyrCysAsp 100
QY 707 TCGTTTGGCAGTGGGACTTTTGCACAACTGCAATTTGTAATTTTGGCAACACTTG 766
```

Db 101 CysPheSerSerGlyAspPheCysAsnSerCysSer-|||||
QY 767 CATCATGATATCAACGGTTAAAGCCATTAGGCGATGCTTGGTAGAATCCAGAGCT 826
Db 117 ArgHisGluLeuGluArgPheLysAlaIleLysAlaCysLeuAspArgProGluAla 136
QY 827 TTCCAGCCAAAATTTGGAGGCGCAATTTGGCAATTTGCAAGCCCGACACACAAAGGG 886
Db 137 PheGlnProLysMetGlyLysGlyArgLeuGlyAlaAlaLysLeuArgHisSerLysGly 156
QY 887 TCGAATCGAGGAGTCAGCTGCCCTGACAGAAATTTACTGCGAGTGTATGAGGCCCAAT 946
Db 157 CysAsnCysLysArgSerGlyCysLeuLysAsnTyrCysGluCysTyrGluAlaLysIle 176
QY 947 ATGTGTTCTTCTATTGCAAAATTCATTGTTGCAAAATTTATGCAAAAGCCCGACGACGA 1006
Db 177 MetCysSerSerIleCysLysCysIleAlaCysLysAsnTyrGluGluSerProGluAsp 196
QY 1007 AAGACACTAATGAGCATGCCAACTACATGCAGACTGGAGTTTGGAGGCGAGCCATTAC 1066
Db 197 LysMetLeuMetSerThrProHisTyrMetGluProGlyAspPheGluSerHisTyr 216
QY 1067 CTGCCACCAACGAATTTTCAGGACTTCCAAAGATTTCAGTCACGATAGGCGGCTTCTCTCA 1126
Db 217 LeuSerProAlaLysPheSerGlyProProLysLeuArgLysAsnArgGlnAlaPheSer 236
QY 1127 TGCATCTCTGGAGGTGGTGGAGGCGACATGCGGCTGCTGCTTCTCAGGAGGAAGAG 1186
Db 237 CysIleSerTrpGluValGluAlaThrCysAlaCysLeuLeuAlaGlnGlyGluGlu 256
QY 1187 GCGAGAAAGAACACCTGCCCTGCCAGACAGATGATCTGGAGGAATTTGGA 1246
Db 257 AlaGluGlnGluHisCysSerProSerLeuAlaGluGlnMetIleLeuGluGluPheGly 276
QY 1247 AGTGCTTATCACAGATTCTCCACTGATGTTAAATCTAAGGAGTTGAAATGGAG 1303
Db 277 ArgCysLeuSerGlnIleLeuHisIleGluPheLysSerLysGlyLeuLysIleGlu 295
RESULT 3
ID LMA5_HUMAN STANDARD; PRT: 3695 AA.
AC O15230; Q9H1P1; Q9WZ47;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LAMA5 OR KIAA0533 OR KIAA1907.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Brown A.J.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahm D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RL "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
RP SEQUENCE OF 197-1934 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
RT DNA Res. 8:179-187(2001).
RN [3]
RP SEQUENCE OF 2051-3695 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RT DNA Res. 5:31-39(1998).
RN [4]
RP SEQUENCE OF 2743-3695 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97415425; PubMed=9271224;
RA Durkin M.E., Loehel F., Mattei M.-G., Gilpin B.J., Albrechtsen R.,
RA Wewer U.M.;
RT "Tissue-specific expression of the human laminin alpha5-chain, and
RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal
RT mouse chromosome 2 near the locus for the ragged (Ra) mutation.";
RL FEBS Lett. 411:296-300(1997).
RN [5]
RP EXPRESSION IN RETINA.
RX MEDLINE=2042761; PubMed=10964957;
RA Libby R.T., Champilaud M.-F., Claudepierre T., Xu Y., Gibbons E.P.,
RA Koch M., Burgess R.E., Hunter D.D., Brunken W.J.;
RT "Laminin expression in adult and developing retinae: evidence of two
RT novel CNS laminins.";
RL J. Neurosci. 20:6517-6528(2000).
CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION AND ORGANIZATION
CC OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -!- SUBUNIT: LAMININ-15 COMPLEX IS AN HETEROTRIMER COMPOSED OF THREE
CC CHAINS (ALPHA-5/BETA-2/GAMMA-3) WHICH ARE BOUND TO EACH OTHER BY
CC DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG
CC AND THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, LUNG, KIDNEY, SKELETAL
CC MUSCLE, PANCREAS, RETINA AND PLACENTA. LITTLE OR NO EXPRESSION IN
CC BRAIN AND LIVER.
CC -!- DOMAIN: DOMAIN G IS GLOBULAR AND IS PART OF THE MAJOR CELL-BINDING
CC SITE LOCATED IN THE LONG ARM OF THE LAMININ HETEROTRIMER.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AL354836; CAC22309.1; ALT_SEQ.
 DR EMBL; AL354836; CAC22310.1; -
 DR EMBL; AB067494; BAB67800.1; -
 DR EMBL; AB011105; BAA25459.1; -
 DR EMBL; Z95636; CAB09137.1; -
 DR HSP; P02468; IKLO.
 DR Genew; HGNC:6485; LMA5.
 DR MIM; 601033; -
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; LamN.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00053; Laminin_EGF; 2.
 DR Pfam; PF00054; Laminin_G; 2.
 DR ProDom; PD002082; LamN; 1.
 DR ProDom; PD003031; Laminin_B; 1.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
 DR PROSITE; PS50025; LAMININ_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 35 POTENTIAL.
 FT CHAIN 36 3695 LAMININ ALPHA-5 CHAIN.
 FT DOMAIN 36 299 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 300 358 LAMININ EGF-LIKE 1.
 FT DOMAIN 359 428 LAMININ EGF-LIKE 2.
 FT DOMAIN 429 474 LAMININ EGF-LIKE 3.
 FT DOMAIN 494 540 LAMININ EGF-LIKE 4.
 FT DOMAIN 541 586 LAMININ EGF-LIKE 5.
 FT DOMAIN 587 631 LAMININ EGF-LIKE 6.
 FT DOMAIN 632 676 LAMININ EGF-LIKE 7.
 FT DOMAIN 677 722 LAMININ EGF-LIKE 8.
 FT DOMAIN 723 775 LAMININ EGF-LIKE 9.
 FT DOMAIN 776 828 LAMININ EGF-LIKE 10.
 FT DOMAIN 829 850 LAMININ EGF-LIKE 11 (INCOMPLETE).
 FT DOMAIN 851 1437 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.
 FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.
 FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.
 FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.
 FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).
 FT DOMAIN 1638 1830 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1831 1863 LAMININ EGF-LIKE 16 (C-TERMINAL).
 FT DOMAIN 1864 1912 LAMININ EGF-LIKE 17.
 FT DOMAIN 1913 1968 LAMININ EGF-LIKE 18.
 FT DOMAIN 1969 2022 LAMININ EGF-LIKE 19.
 FT DOMAIN 2023 2069 LAMININ EGF-LIKE 20.
 FT DOMAIN 2070 2116 LAMININ EGF-LIKE 21.
 FT DOMAIN 2117 2166 LAMININ EGF-LIKE 22.
 FT DOMAIN 2167 2735 DOMAIN II AND I.
 FT DOMAIN 2736 2929 LAMININ G-LIKE 1.
 FT DOMAIN 2941 3115 LAMININ G-LIKE 2.
 FT DOMAIN 3124 3292 LAMININ G-LIKE 3.
 FT DOMAIN 3340 3513 LAMININ G-LIKE 4.
 FT DOMAIN 3520 3692 LAMININ G-LIKE 5.
 FT DOMAIN 2203 2221 COILED COIL (POTENTIAL).
 FT DOMAIN 2335 2466 COILED COIL (POTENTIAL).
 FT DOMAIN 2510 2670 COILED COIL (POTENTIAL).
 FT SITE 1722 1724 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1838 1840 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 300 309 BY SIMILARITY.
 FT DISULFID 302 322 BY SIMILARITY.
 FT DISULFID 324 333 BY SIMILARITY.
 FT DISULFID 336 356 BY SIMILARITY.
 FT DISULFID 359 368 BY SIMILARITY.
 FT DISULFID 361 393 BY SIMILARITY.
 FT DISULFID 396 405 BY SIMILARITY.
 FT DISULFID 408 426 BY SIMILARITY.
 FT DISULFID 429 440 BY SIMILARITY.
 FT DISULFID 441 447 BY SIMILARITY.

FT DISULFID 449 458 BY SIMILARITY.
 FT DISULFID 461 471 BY SIMILARITY.
 FT DISULFID 494 506 BY SIMILARITY.
 FT DISULFID 496 515 BY SIMILARITY.
 FT DISULFID 517 526 BY SIMILARITY.
 FT DISULFID 529 538 BY SIMILARITY.
 FT DISULFID 541 553 BY SIMILARITY.
 FT DISULFID 543 560 BY SIMILARITY.
 FT DISULFID 562 571 BY SIMILARITY.
 FT DISULFID 574 584 BY SIMILARITY.
 FT DISULFID 587 599 BY SIMILARITY.
 FT DISULFID 589 605 BY SIMILARITY.
 FT DISULFID 607 616 BY SIMILARITY.
 FT DISULFID 619 629 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT DISULFID 634 650 BY SIMILARITY.
 FT DISULFID 652 661 BY SIMILARITY.
 FT DISULFID 664 674 BY SIMILARITY.
 FT DISULFID 677 689 BY SIMILARITY.
 FT DISULFID 679 696 BY SIMILARITY.
 FT DISULFID 698 707 BY SIMILARITY.
 FT DISULFID 710 720 BY SIMILARITY.

Alignment Scores:

Pred. No.: 0.0298 Length: 3695
 Score: 130.50 Matches: 111
 Percent Similarity: 31.14% Conservative: 55
 Best Local Similarity: 20.83% Mismatches: 198
 Query Match: 3.41% Indels: 170
 DB: 1 Gaps: 28

US-09-743-237-3 (1-2134) x LMA5_HUMAN (1-3695)

QY 18 AAGTCGCGGGGGCGGACAGCGCGGGGAG-----CTCCTCGGGGAGTAC 65
 DB 83 LysLeuValGlyGlyProValAlaGlyAspProAsnGlnThrIleGlnGlyGluTyr 102
 QY 66 CCGGGGATC-----CCAGAGCTCAGCGCGCTGGAG 95
 DB 103 CysAspIleCysThrAlaAlaAsnSerAsnLysAlaHisProAlaSerAsnAlaIleasp 122
 QY 96 GACGTC-----GCGCTCTCGAGCGCCGCGAGCGCGCGCTGC 134
 DB 123 GlyThrGluArgTrpTrpGlnSerProProLeuSerArgGlyLeuGluTyrAsnGluVal 142
 QY 135 AACGTG-----CACTTCCTGCTCGCTCGCTACCCGCGCAC 170
 DB 143 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 162
 QY 171 CGACGCGCGGGGTGTTTGGCCCTGGCGGCGCTGCTGCGGAAGAGAGCTCCACCC 230
 DB 163 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerMetAspPhe 178
 QY 231 GGGCGTCGCGATGATCCAGTTGAAATCAAGTAAGCAGGTGGTACTACTACAAGTAATA 290
 DB 179 GlyArgThrTyrGlnPro-----TrpGlnPhePhe 188
 QY 291 ATCCGGAAGAGCAACTTTCGAGAACTCTTGTCTCAGGAATCCTGTGCAAGTTCCCAT 350
 DB 189 AlaSerSerLysArgAspCys-----LeuGluArgPhe 199
 QY 351 GGTCCCGAGGAACATA-----GAGGATGCTCTGCTGCTCTTAAAGAAA 394
 DB 200 GlyProGlnThrLeuGluArgIleThrArgAspAspAlaIleCysThrThrGluTyr 219
 QY 395 GATTCCAAACCAATGGTGATATGCCAATTAAGAGGGGGCACACAAATGCTATGATAGAC 454
 DB 220 SerArg-----IleValProLeuGluAsnGlyGluIleValValSerLeuVal 235
 QY 455 AATCTAGACACAGAGCAACTAAAGCAGCTCCATTGGTTCTCAGTACGATCAAAAT 514
 DB 236 AsnGlyArgProGlyAlaMet----- 242


```
QY 926 GAGTGCATATGAG-----GCCCAAAATATGTTCTTCTTATTTGCAAA 967
||||| |||
Db 293 GluCysLysGluValLysCysAsnGlyGlnValPheCysLysAspSerCysSer 312
||| |||
QY 968 TGCATTGGTGCACAAATATGAGAA-----ACCCAGACAGAAAGACA 1012
||||| |||
Db 313 CysVal---CysProGlyGlyAspLysThrCysThrAlaProGlnValTyAsp 331
||||| |||
QY 1013 CTAATGAGCAGCCAACTACATCAGACTGGAGGTTTGAAGCAGCCATACCTGCCA 1072
||| |||
Db 332 GlyVal-AlaCysSerCysCysProValAsnMetGlnLysProAlaAspGlyCysPr 351
||||| |||
QY 1073 CCACGAAATTTCCAGGATCTCCAGATC-----AGTCACGATAG 1113
||| |||
Db 351 oArgProGlnLysTrpAspLysGluGluCysArgCysGluCysProValLysHisAspCy 371
||| |||
QY 1114 CGCGCTTCCTCATGCTCTCTGGAGGTGGTGAGCGCCATGCGCTGCTGCTGTC 1173
||| |||
Db 371 sLysAsnGlyLysVal-----TrpAsp-----GluThrIleCysGlnCysIleCysPr 387
||| |||
QY 1174 TCAG-----GGAGAAGAGAGCCGAG----- 1192
||| |||
Db 387 oArgAspAlaProValCysThrAlaGlyLysGluArgCysGlyGluSerCysGluCysLy 407
||| |||
QY 1193 -----AAAGAACACTGCTCCAAG----- 1210
||| |||
Db 407 sCysIleAsnArgGluProLysGluGlyCysAlaLysProIleValTrpAsnGluAsnTh 427
||| |||
QY 1211 -----TGCTGGCAGACAGATGATCTG-----GA 1236
||| |||
Db 427 rCysLysCysValCysProAlaAspLysGlnMetSerProGlyGlyCysGlySerGlyLy 447
||| |||
QY 1237 GGAATTTGGAAGTGCTATACAGATCTCCACACGAGTTTAAACTAAG---GCATT 1293
||| |||
Db 447 sSerPheAsnLysLeuThrCysGlnCysGluCysAspGlnSerAlaSerLysCysGlyLe 467
||| |||
QY 1294 GAAATG-----GAGTAGAGTATAAGTGTGAATGATGCTGTGTTGTTGCTAGTCTAG 1347
||| |||
Db 467 uLysArgTrpAsnAlaAspThrCysLysCysGluCys----- 479
||| |||
QY 1348 AAATCTCTAGTTTGAAGAAGATGTTTAGGGGAACATGAGGCTGCTGCAGCACAACC 1407
||| |||
Db 480 -----GlnPr 481
||| |||
QY 1408 AGGCTCCCTGCTCATCTGCTGGCCGAGGAGTTTACTCA----- 1445
||| |||
Db 481 oGlyMetProGlu---GlyCysGlyLys-GlnThrTrpIleSerAspLysCysLysC 500
||| |||
QY 1446 --GAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTCTGAGGAGGTGCAGGCGCTGA 1503
||| |||
Db 500 ysGluCysSerProThrIleThrCysGlnAlaProGlnIleLeuAspLeuAsnThrCysG 520
||| |||
QY 1504 GCATTGTTGT-----CTGCCGAGAGGAGAGCTGGGTTCCCATAGTCTCT 1551
||| |||
Db 520 luCysLysCysProValAsnMetLeuAlaGlnLysCysLysSerProArgGlnT 540
||| |||
QY 1552 GG-----GAGAGTCTCTGAGGCGCGGAGGCGAGAGCCCTGCGGAGAGCT 1602
||| |||
Db 540 rpThrAspSerLysCysLeuCysGlyCysSerThrThrProAlaThrCysGluGlyLysG 560
||| |||
QY 1603 CACTCTGCTGACACTTCTCTCAGAGAAATGTGCTCTGAGGCG 1646
||| |||
Db 560 lnThrTrpCysGlyGluAlaCysGlnCysIleCysProGlyGly 574
||| |||
RESULT 5
AHM2_ARATH STANDARD; PRT; 1172 AA.
AC 064474;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Potential cadmium/zinc-transporting ATPase 2 (EC 3.6.3.3)
```

(EC 3.6.3.3).
HMA4 OR AT2G19110 OR T20K24.12.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia; TISSUE=Root;
RC Krijger G.C., Baccarini P.J., Williams L.E.;
RA "Identification of AthMA4, a putative heavy metal P-type ATPase in
RT Arabidopsis.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RC MEDLINE=20083487; PubMed=10617197;
RX Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rinning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umeyam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RL Nature 402:761-768 (1999).
CC -|- FUNCTION: INVOLVED IN CADMIUM/ZINC TRANSPORT (POTENTIAL).
CC -|- CATALYTIC ACTIVITY: ATP + H(2)O + Cd(2+)(In) = ADP + phosphate +
CC Cd(2+)(Out).
CC -|- CATALYTIC ACTIVITY: ATP + H(2)O + Zn(2+)(In) = ADP + phosphate +
CC Zn(2+)(Out).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
(E1-E2 ATPASES). SUBFAMILY 1B.
CC -|- SIMILARITY: CONTAINS 1 HMA DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL: AJ297264; CAC19544.1; -;
DR EMBL: AC002392; AAD12041.1; -;
DR InterPro: IPR001757; ATPase_E1-E2.
DR InterPro: IPR001756; Cu_ATPase.
DR InterPro: IPR001934; HeavyM_e_transpt.
DR InterPro: IPR001454; Hlgpase/hydrlase.
DR Pfam: PF00122; E1-E2_ATPase; 1.
DR Pfam: PF00403; HMA; 1.
DR Pfam: PF00702; Hydrolase; 1.
DR PRINTS: PR00119; CATATPASE.
DR PRINTS: PR00943; CATATPASE.
DR PROSITE: PS00154; ATPase_E1_E2; 1.
DR PROSITE: PS01047; HMA_1; FALSE_NEG.
DR PROSITE: PS0846; HMA_2; 1.
KW Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Magnesium;
Zinc; Cadmium; Metal-binding; Multigene family.
FT DOMAIN 1 93
FT TRANSMEM 94 115
FT DOMAIN 116 118
FT TRANSMEM 119 138
FT DOMAIN 139 145
FT TRANSMEM 146 166
FT DOMAIN 167 167
FT TRANSMEM 168 188
FT DOMAIN 189 314
FT TRANSMEM 315 337
POTENTIAL.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
POTENTIAL.

FT DOMAIN 338 345 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 346 363 POTENTIAL.
 FT DOMAIN 364 656 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 657 676 POTENTIAL.
 FT DOMAIN 677 680 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 681 700 POTENTIAL.
 FT DOMAIN 701 1172 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1172 1172 HMA.
 FT MOD_RES 1157 1169 POLY-HIS.
 FT METAL 401 401 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 601 601 MAGNESIUM (BY SIMILARITY).
 FT METAL 605 605 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 1172 AA; 127208 MW; 1E913B9450443440 CRC64;

Alignment Scores:
 Pred. No.: 0.17 Length: 1172
 Score: 119.50 Matches: 110
 Percent Similarity: 29.77% Conservative: 57
 Best Local Similarity: 19.57% Mismatches: 184
 Query Match: 3.12% Indels: 211
 DB: 1 Gaps: 23

US-09-743-237-3 (1-2134) x AHW2_ARATH (1-1172)

QY 235 GTCCGATCATCCAGTGAATCAAGTAAAGCAGGTGGTACTACTACAAAGTAATAATCC 294
 Db 639 IleArgArgIleProGlnAlaValLysLeuAlaArg-----Arg 651
 QY 295 GGAAGAAGCAACTTTCAGAAATCTCTGCTCAGGNATCCTGTGCAAGTCCCATGGTC 354
 Db 652 AlaArgArgLysValValGluAsnValCysLeuSerIleLeuLys----- 667
 QY 355 CCAGGAACATAGAGGATGCTCTGCTCTTAAGAAAGATTCAACCCCAATGGTG-- 412
 Db 668 -----AlaGlyIleLeuAlaLeuAlaPheAlaGlyHisProLeuIleTr 582
 QY 413 ----ATATGCCAATTTGAAGGGGCGACAAATGCTATGATATAGCAATTCAGA----- 463
 Db 682 pAlaAlaValLeuValAspValGlyThrCysLeuLeuValIlePheAsnSerMetLeuLe 702
 QY 464 -ACAAGAGAACTAAAGCACTCCATTTGGTTCTCTCAGTATCAAGATCAAAATAATTATCT 522
 Db 702 uLeuArgGluLysLysLysIleGlyAsnLysLysCysTyrArg----- 716
 QY 523 ACAGTCAGATGCTCTAAACCAATGACTCTTTAGTAGGAGATTTCGCCAGCATCAAC 582
 Db 717 -----AlaSerThrSe 720
 QY 583 AAAATAAATCATTTACACAACTTGAGGAGCCTTACCATCGGTAGTCAAC----- 637
 Db 720 rLysLeuAsn-----GlyArgLysLeuGluGlyAspAspTyrValValAspLeuGlu 738
 QY 638 ---GGGTCTGCTTCCCTCGGGATCACTCTTCCAGGACCACCAAAATAACTTTGGC 693
 Db 738 uAlaGlyLeuLeuThrLysSerGlyAsn----- 747
 QY 694 TGGGTACTGTGACTGCTTTGCCAGTGGGACTTTTGCAACAACTGCAAT----- 742
 Db 748 -GlyGlnCysLysSerSerCysGlyAspLysLysAsnGlnGluAsnValValMetMe 767
 QY 743 -----TGT----- 745
 Db 767 tLysProSerSerLysThrSerSerAspHisSerHisProGlyCysGlyAspLysLys 787
 QY 746 -----AATAATGTTTC----- 757
 Db 787 sGluGluLysValLysProLeuValLysAspGlyCysSerSerGluLysThrArgLysSe 807
 QY 758 -----AACAACTTCATCATGATAT 777
 Db 807 rGluGlyAspMetValSerLeuSerSerCysLysLysSerSerHisValLysHisAspLe 827
 QY 778 TGAACGGTTTAAAGCATTAAGGCATGCTTGGTAGAAATCCAGAAGCTTTCCAGCAAA 837

Db 827 u---LysMetLysGlyGlySerGlyCysAlaSerLysAsnGluLysGlyLysGluVa 846
 QY 838 AATTGGGAAGGCCCAATTGGGCAATGTCAAGCCCAAGCACACAAGGTTGCAACTGCAG 897
 Db 846 lValAlaLysSerCysCysGluLysProLysGlnValGluSerValGlyAspCysLys 866
 QY 898 GAGGTGAGCTGCTGCTGAAG----- 916
 Db 866 sSerGlyHisCysGluLysLysGlnAlaGluAspIleValProValGlnIleIl 886
 QY 917 ----AATTACTCGAGTGTATGAGGCCCAAAATTTATGTTCTTCTATTTCGAAATGCAT 972
 Db 886 eGlyHisAlaLeuThrHisValGluIleGluLeuGlnThrLysGluThrCysLysThrSe 906
 QY 973 TGGTTGCCAAAATTTAAGAAAGCCCAAGAACGACAACTAATGAGCATGCCAAAC-- 1030
 Db 906 rCysCysAspSerLysGluLysValLysGluThrGlyLeuLeuLeuSerSerGluAsnTh 926
 QY 1031 ----TACATGCAGACTGGAGTTTG-----GAAGGCAGCCA 1062
 Db 926 rProTyrLeuGluLysGlyValLeuIleLysAspGluGlyAsnCysLysSerGlySerGl 946
 QY 1063 TTACCTGCCCAACCAAGAAATTTTCAGGACTTCCAAAGATTTCAGTCAGATGAGCGCCTTC 1122
 Db 946 uAsnMetGlyThrValLysGlnSerCysHisGluLysGlyCysSerAspGluLysGlnTh 966
 QY 1123 CTCATGCATCTCTCTGGGAGGTGGTGAGGCCACATGCCGCTGCTGCTGCTCAGGGAGA 1182
 Db 966 rGlyGluIleThr-----LeuAlaSerGluGl 975
 QY 1183 AGAGGCCGAGAAGAACACTGCTGCTCAAG-----TGCCTGGCAGAGCAGCATCTCTGA 1236
 Db 975 uGluThrAspAspGlnAspCysSerSerGlyCysCysValAsnGluGlyThrValLysGl 995
 QY 1237 GGAATTTGGAAGGTGCTTATCAGATTCACAGATTCTCCACACTGATTTAAATCTAAGGATTTAA 1296
 Db 995 nSerPheAspGluLysLysHisSerValLeu-----ValGluLysGluGlyLeuAs 1012
 QY 1297 AATGAGTAGATATAAAGTGTGAATGCATGTTGATTTTGTCTTAGTCTAGAAATCTCTA 1356
 Db 1012 pMetGluThrGlyPheCysCysAspAlaLysLeuValCys----- 1025
 QY 1357 GTTTAGAAAGGATGTTTAGGGGAACATGAGGCTGCTGTCAGCAACAACAGGCTCCCC 1416
 Db 1025 ----- 1025
 QY 1417 TCATCTCCCTGGCCCAAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCC 1476
 Db 1026 -----CysGlyAsnThr----- 1029
 QY 1477 CCTTTTCTGAGAGGTGCTGCTGCTGAGCATTTGT---TTGCTGGCCCAAGAGAGAGAC 1533
 Db 1030 -----GluGlyGluValLysGluGlnCysArgLeuGluIleLysLysGluGlu-- 1045
 QY 1534 TTGGGTTCCCATAGTCTGGGAGAGTGTCTGCAGGCGCGGAGGAGCAGCAGGCGCCG 1593
 Db 1046 -----HisCysLysSerGlyCysGlyGluGluIleGlnTh 1058
 QY 1594 CGGAGAGTCACTCTGGTGCAGTCTTCTCTCTCAGAGAAATGT-----TGCTTGGAGGCT 1647
 Db 1058 rGlyGluIleThrLeuValSerGluGluGlu-ThrGluSerThrAsnCysSerThrGlyC 1078
 QY 1648 GC 1649
 Db 1078 ys 1078
 RESULT 6
 POLG_HCVJA
 ID POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)

Db	1284	ThrIleThrThrGlyGlyPro-----	-----IleThrTyrSerThr	1299
QY	335	TGTTGCAAGTTCCTCATGGTCCCGAGGAAGCTAGAGGATGCTCTGCTGCTCTCTTAAAGAA	394	
Db	1296	TyrCysLysPhe-----	-----LeuAlaAspGlyGlyCysSerGlyGlyAlaTyr	1310
QY	395	GATTCCAAACCCAAATGCTGATATGCCAAATGGAAGGGGCACACAAATGCTATGATATA---	451	
Db	1311	Asp-----IleIleIleCysaspGluCysHisSerThrAspSerThrThrIleLeu	1327	
QY	452	-----CACAAATCTAGACAACAGAGAACATAAAGACACTCCATTGGTTGGTT	493	
Db	1328	GlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValValLeuAla	1347	
QY	494	-----CCTCAGTATCAAGATCAAAATAAAT	517	
Db	1348	ThrAlaThrProProGlySerIleThrValProHisProAsnIleGluGluValAlaLeu	1367	
QY	518	TATCTACATGCAGATGCTCCCT-----	562	
Db	1368	SerAsnThrGlyGluIleProPheTyrGlyLysAlaIleProIleGluAlaIleLysGly	1387	
QY	563	AGA-----TTTTTGCACGATCAACAAATAATAATCTCATTTACACACAACCTT	610	
Db	1388	GlyArgHisIleLeuPheCysHisSerLysLysCysAspGluLeuAlaAlaLeuAla	1407	
QY	611	GAGGGA-----GCCTTACCATCTCGGTA-----	652	
Db	1408	ThrGlyLeuGlyLeuAsnAlaValAlaTyrTyrArgGlyLeuAspValSerValIlePro	1427	
QY	653	-----TCGGGATCAACTTCCAGGACACCAAAATAACTTTGGCTGGGTACTGTGACTGC	709	
Db	1428	ThrSerGlyAspValValValAlaThrAspAlaLeuMetThrGlyPhe-----	1444	
QY	710	TTTGGCCAGTGGGACTTTTGGCAACACTGCCAATTTGTAATAATTTGTTGCCACAACCTTG-	766	
Db	1445	-----ThrGlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrVal	1462	
QY	767	-----CATCATGATATTGAACGGTTTAAAGCCATTAAAGCCATTAAAGGCATGCTTT	808	
Db	1463	AspPheSerLeuAspProThrPheThrIleGluThrThrThrLeuProGlnAspAlaVal	1482	
QY	809	GGTAGAAATCCAGAAGCTTTCCAGGCCAAAATTTGGGAAGGCCAAATTTGGG-----	859	
Db	1483	SerArgAlaGlnArg-----ArgGlyArgThrGlyArgGlyArgSerGlyIleTyrArg	1500	
QY	860	AATGTCAGCCCCGACACACAAAGGGTCCAACTCCAGGAGGTGAGCTGCCTCGCTCAAGAAT	919	
Db	1501	PheValThrProGlyGlu-----ArgProSerGlyMetPheAspSer	1514	
QY	920	-----TACTGCGAGTGCCTATGAGGCCAACAAATTATGTGTTCTTCTATTGTCAAAATGCATT	973	
Db	1515	SerValLeuCysGluCysTyrAspAla-----	1523	
QY	974	GTTTGCAAAATATATGAGAAAGCCCGAAGAACACTAATGAGCACTGCCCAAACTAC	1033	
Db	1524	GlyCysAlaTrpTyrGluLeuThrPro--AlaGluThrSerValArgLeuArgAlaTyr	1542	
QY	1034	ATCCACACTGGAGTTTGAA-----	1054	
Db	1543	LeuAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluSerValPhe	1562	
QY	1055	-----GGCAGCATTAAGTCCGCCAACCAAGAAATTTTTCAGGA-----	1090	
Db	1563	ThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsn	1582	
QY	1091	CTTCCAAGATTTCAGTC-----	1138	
Db	1583	LeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProPr	1602	
QY	1139	GAGCTGG-----	1189	
Db	1602	oserTrpaspGlnMetTrpLys-----	1617	

FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1088 1088 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1252 1252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1351 1351 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1502 1502 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1801 AA: 196473 MW; 97AEF32F8F31FA75 CRC64;

Alignment Scores:
 Pred. No.: 0.222 Length: 1801
 Score: 118.50 Matches: 110
 Percent Similarity: 27.02% Conservative: 54
 Best Local Similarity: 18.12% Mismatches: 193
 Query Match: 3.10% Indels: 250
 DB: 1 Gaps: 29

US-09-743-237-3 (1-2134) x LMB2_RAT (1-1801)

QY 76 CAGAGCTCAGCGGCTGGAGGAGCTCGCGCTCTGCGAGGCCCGCAGCCGCCCTGCA 135
 Db 633 GlnValProGluGlnTrpAlaGluLeuGluValValGlnArgProGlyProValSer 652
 QY 136 ACGTGCACTTCCTGCTCTGCTACCGCGCACCCGCGCGGGTGTTCGCCCT 195
 Db 653 AlaHisSerProCysGlyHisValLeuProArg- 663
 QY 196 GGGCGCGCTGGGCTCGGAGGAGGCTCCACCGCGGCTCGCGATGATC- 246
 Db 664 -----AspAspArgIleGlnGlyMetLeuHisProAsnThrArgValLeuValPhePro 681
 QY 247 ---CCAGTGAATCAAGGTAACGAGTGTACTACTACAG- 291
 Db 682 ArgProValCysLeuGluProGlyLeuSerTyrLysLeuLysLeuLysLeuThrGlyThr 701
 QY 292 TCCGGAAGAGCAAC---TTTGAGAAATCTTCTGCTCAGGAATCCTGTG- 345
 Db 702 GlyGlyArgAlaHisProGluThrProTyrSerGlyLeuLeuLeuLeuLeuLeuLeu 721
 QY 346 CCCATGGTCCCGAACTAGAGGAGCTCCTGCTGCTTCTTTAAGAAAGATTCCAACC 405
 Db 722 ValLeuGlnProHisVal-----LeuMetLeu-----GluMetPheSerGly 735
 QY 406 AATGGTGATATGCCAATTTGAAGGGGCGACAAATGCTATGATATAGACAAATTCAGAAC 465
 Db 736 GlyAspAlaAlaAlaLeuGluArgArgThrPheGluArgTyrArg----- 751
 QY 466 AAGAGACAATAAAGCACTCCATTGGTTCTCTCAGTATCAGATCAAAATATATCTACA 525
 Db 751 ----- 751
 QY 526 GTCAGATGTCCCTAAACCAATGACTGCTTTAGTAGGAGATTTTGGCAGCATCAACAA 585
 Db 752 -----CysHisGluGluGlyLeu-MetProSerLysThrPr 763
 QY 586 ATTAATATCATTAACACAACTTACAGGAGCCTTACCATGGTAGTCAACGGGTCTGC 645
 Db 763 oLeuSer-----GluAlaCysValProLeuLeuIleSerAlaSerSe 777
 QY 646 TTTCCCTCGGGATCAACTTTCAGGACCACCAAAATAACTTTGGCTGGGTACTGTGA 705
 Db 777 iLeuValTyrAsnGlyAlaLeuPro-----CysG1 787
 QY 706 CTGCTTTGGCAGTGGGAGCTTTTGGCAACACTGCAATTTGTAATAATTTGTCACAACTT 765
 Db 787 nCysAspProGlnGlySerLeuSerSerGluCysAsnProHisGly----- 802
 QY 766 GCATCATGATATGAAGGTTTAAAGCCATTAGGCATCTCTTGGTGTAGAAATCCAGAAC 825
 Db 803 -----GlyGlnCysArgCysLysProGlyValValGlyArgArgCysAspAl 818
 QY 826 TTTCCAGCAAAAATTTGGGAGGGCAATTTGGCATGTCAAGCCCGCCACCAACAAAGG 885
 Db 818 a-----CysAlaThrGlyTyrTyrGly-----PheGlyProAlaGlyCysGlnAl 833

RESULT 9

ZAN_MOUSE

ID ZAN_MOUSE

AC 088799; 008647;

STANDARD;

PRT; 5376 AA.

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-Testis;
RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT "Species diversity in the structure of zonadhesin, a sperm-specific,
RT membrane protein containing multiple cell adhesion molecule-like
RT domains.";
RL J. Biol. Chem. 273:3415-3421(1998).
RN [2]
RN SEQUENCE OF 4864-5376 FROM N.A.
RC TISSUE-Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD.
CC -1- TISSUE SPECIFICITY: IN TESTIS: PRIMARILY IN HAPLOID SPERMATIDS.
CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOEA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -1- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
CC -1- SIMILARITY: CONTAINS 25 VWFD DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U97068; AAC26680.1; -;
CC EMBL: U83190; AAC53125.1; -;
CC MGD: MGI:106656; ZAN.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR003645; FOLN.
CC InterPro: IPR000998; MAM domain.
CC InterPro: IPR002919; TIL_Cysrich.
CC InterPro: IPR003328; TILA_Cysrich.
CC InterPro: IPR001007; VWF_C.
CC InterPro: IPR001846; VWF_D.
CC Pfam: PF00094; vwd; 4.
CC Pfam: PF00629; MAM; 3.
CC Pfam: PF01826; TIL; 25.
CC Pfam: PF02345; TILA; 25.
CC SMART: SM00181; EGF; 2.
CC SMART: SM00274; FOLN; 11.
CC SMART: SM00137; MAM; 2.
CC SMART: SM00214; VWC; 17.
CC SMART: SM00216; VWD; 4.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01186; EGF_2; 18.

DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00060; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 5376 ZONADHESIN.
FT DOMAIN 18 5310 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5311 5337 POTENTIAL.
FT DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 45 210 MAM 1.
FT DOMAIN 215 374 MAM 2.
FT DOMAIN 377 542 MAM 3.
FT DOMAIN 547 1170 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT (MUCIN-LIKE DOMAIN).
FT DOMAIN 1171 1280 VWFD 1 (PARTIAL).
FT DOMAIN 1281 1669 VWFD 2.
FT DOMAIN 1670 2056 VWFD 3.
FT DOMAIN 2057 2459 VWFD 4.
FT DOMAIN 2460 2579 VWFD 5 (PARTIAL).
FT DOMAIN 2580 2699 VWFD 6 (PARTIAL).
FT DOMAIN 2700 2819 VWFD 7 (PARTIAL).
FT DOMAIN 2820 2939 VWFD 8 (PARTIAL).
FT DOMAIN 2940 3059 VWFD 9 (PARTIAL).
FT DOMAIN 3060 3179 VWFD 10 (PARTIAL).
FT DOMAIN 3180 3299 VWFD 11 (PARTIAL).
FT DOMAIN 3300 3416 VWFD 12 (PARTIAL).
FT DOMAIN 3417 3536 VWFD 13 (PARTIAL).
FT DOMAIN 3537 3656 VWFD 14 (PARTIAL).
FT DOMAIN 3657 3776 VWFD 15 (PARTIAL).
FT DOMAIN 3777 3892 VWFD 16 (PARTIAL).
FT DOMAIN 3893 4029 VWFD 17 (PARTIAL).
FT DOMAIN 4029 4148 VWFD 18 (PARTIAL).
FT DOMAIN 4149 4263 VWFD 19 (PARTIAL).
FT DOMAIN 4264 4283 VWFD 20 (PARTIAL).
FT DOMAIN 4504 4623 VWFD 21 (PARTIAL).
FT DOMAIN 4624 4743 VWFD 22 (PARTIAL).
FT DOMAIN 4744 4863 VWFD 23 (PARTIAL).
FT DOMAIN 4864 5261 VWFD 24 (PARTIAL).
FT DOMAIN 5259 5295 EGF-LIKE.
FT DISULFID 5263 5274 BY SIMILARITY.
FT DISULFID 5268 5283 BY SIMILARITY.
FT DISULFID 5285 5294 BY SIMILARITY.
FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1216 1216 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1314 1314 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1814 1814 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1933 1933 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2028 2028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2111 2111 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2142 2142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2332 2332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2575 2575 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2692 2692 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2812 2812 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3052 3052 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3065 3065 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3144 3144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3172 3172 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3288 3288 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3292 3292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3782 3782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4005 4005 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4136 4136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4243 4243 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4254 4254 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4335 4335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4376 4376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4586 4586 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 5136 5136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5252 5252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 5376 AA; 579908 MW; 0B44DB77DF2A2620 CRC64;

Alignment Scores:
 Pred. No.: 0.278 Length: 5376
 Score: 118.50 Matches: 137
 Percent Similarity: 29.39% Conservative: 64
 Best Local Similarity: 20.03% Mismatches: 232
 Query Match: 3.10% Indels: 251
 DB: 1 Gaps: 41

US-09-743-237-3 (1-2134) x ZAN_MOUSE (1-5376)

QY 16 CGAAGTCGGGGGGGCGGACGAGCGGGGAGCTCTCGGGGAGTACCCCGGATCC 75
 Db 3478 LysAspAlaGlnGlyAlaLeuIleProAlaGlyLysThrTrpThrSerProGlyCysThr 3497
 QY 76 CAGAGTCAGCGCGCTGGAGGAGCTCGGCTCTCGAGGCCCGCGAGCCCGCCCTGCA 135
 Db 3498 GlnSerCysAlaCysMetGlyGlyAlaValGlnCysGlnSerSerGlnCysProGly 3517
 QY 136 ACG---TGC-----ACTTCCTGCTCTCGC-----TGC 159
 Db 3518 ThrTyrcysLysAspGlnGluAspGlyAsnSerAsnCysAlaLysIleThrLeuGlnCys 3537
 QY 160 TACCCGCGCACCGCGCGGGTGTGTTTCCCTCGGGCGCCT-----204
 Db 3538 ProAlaHisSerLeuPheThrAsnCysLeuProCysLeuProSerCysLeuAspPro 3557
 QY 205 ---GGTCTCGCAAGAGGCTCCACCGCGGGCTCGGATGATCCAGTTGAATCAAG 261
 Db 3558 AspGlyLeuCysLysGlyAlaSer-----3565
 QY 262 GTAAGCAGGTGGTACTACTACAAGTAATAATCCGGAAGAAGCACTTTCAGAACTCTCT 321
 Db 3565 -----3565
 QY 322 TGCTCAGGAATCCTGTTGCAAGTTCCTCGTCCCGCAAGT-----AGAGGATCCCTC 375
 Db 3566 -----ProLysValProSerThrCysLysGluGlyCysIle 3577
 QY 376 CTGC-----TGTCTCTTAAG-----391
 Db 3578 -CysGlnSerGlyTyrrValLeuSerAsnAsnLysCysLeuLeuArgAsnArgCysGlyCy 3597
 QY 392 -AAAGATTCCAAACCAATGCTGATA-----TGTCTCTTAAG-----415
 Db 3597 sLysAspAlaHisGlyAlaLeuIleProGluAspLysThrTrpValSerArgGlyCysTh 3617
 QY 416 -----TGCCAAATGAAAGGGGCACA---CAAATGCTA-----445
 Db 3617 rGlnSerCysValCysThrGlyGlySerIleGlnCysLeuSerSerGlnCysProGly 3637
 QY 446 -----TGTATAGACAATCTAGACACAGAGACTAAAGCAGCTCCATTTGGTCTCA 498
 Db 3637 yAlaTyrcysLysAspAsn---GluAspGlySerSerAsnCysAlaArgIleProGly 3656
 QY 499 GTATCAAGATCAAAATAATATATACAGTCAGATGCTCCCTAAACCAATGACTGCTTTACT 558
 Db 3656 nCysProAlaAsnSerHisTyrrThrAspCysPheProCysProProSerCysSerAs 3676
 QY 559 AGGAGATTTTGGCCAGCATCAACAAAATAATATATATACACACAACTTGAGGAGC 618
 Db 3676 pProGluGlyHisCysGluAlaSerGlyProArgValLeuSerThrCysArgGluGlyCy 3696
 QY 619 CTTA-----CCATCGGTA-----631
 Db 3696 sLeuCysAsnProGlyPheValLeuAspArgAspLysCysValProArgValGluCysG 3716
 QY 632 -----GTCAAGGGTGTCTTCCCTCGGGATCAACTCTTCCAGGACCA-----676
 Db 3716 yCysLysAspAlaGlnGlyAlaLeuIleProSerGlyLysThrTrpThrSerProGlyCy 3736

QY 677 -----CCAAAAATAACTTTGGCTGGTACTGTGACTGTGCTTGGCCAGT-----718
 Db 3736 sThrGlnSerCysAlaCysMetGlyGlyValValGlnCysGlnSerSerGlnCysProPr 3756
 QY 719 -GGGAGCTTTTGC-----AACAACTCCAATTTGAATAATTTGTTGCAACAA 762
 Db 3756 oGlyThrTyrcysLysAspAsnGluAspGlyAsnSerAsnCysAlaLysIleThrLeuG 3776
 QY 763 CTTCATCATGATATTGAACGGGTTTAAAGCCATTAAAGCATGCTTGGTGAATAACAGA 822
 Db 3776 nCysProThrHisSerAsnTyrrThrAspCysLeuProPheCysLeu-----Prose 3793
 QY 823 AGCTTTCCAGCCAAAAATTGGGAAGGCCAAATTGGCAATGTCAAGCCCCAGCAC-----877
 Db 3793 rCysLeuAspPro-----SerAlaLeuCysGlyGlyThrSerProLysGlyProse 3810
 QY 878 -----ACAAGAGGTGCAACTGCAGG-----AGTCTAGGCTGC--910
 Db 3810 rThrCysLysGluGlyCysValCysGlnProGlyTyrrValLeuAspLysAspLysCys 3830
 QY 911 -CTGAGAATTACTCGGAGTGTCTATGAGCCCAA-----943
 Db 3830 eLeuLysIleGluCysGlyCysArgAspThrGlnGlyAlaValIleProAlaGlyLysTh 3850
 QY 944 -----ATTATGTGTTTCTTATTTCGAAATGC-----ATTGTTTGCAC 981
 Db 3850 rTrpLeuSerThrGlyCysIleGlnSerCysAlaCysValGluGlyThrIleGlnCysG 3870
 QY 982 AAATTATGAAGAAAGCCAGAA-----CGAAGACACT 1014
 Db 3870 nAsnPheGlnCysProProGlyThrTyrrCysAsnHisAsnAsnCysAlaLysIlePr 3890
 QY 1015 AATGAGCATGCAACACTACATCAGACTCGAGGTTTGAAGGAGCAGCATTAATCTGCCACC 1074
 Db 3890 oLeuGlnCysProAlaHis-----SerHisPheThr-----3900
 QY 1075 AACGAAATTTTCAGGACTTCCAAGATTCACTCAGATAGCGGCTTCTCATGCTATCTC 1134
 Db 3901 -----SerCysLeuProSerCys-----ProProSerCysAlaAs 3912
 QY 1135 CTGGGAGGTGGTGGAGGCCACATGCGCTGCTGCTGCTCAGGAGAGAGAGCCGAGAA 1194
 Db 3912 n-----LeuAspGlySerCys-----GluGlnThrSerProLysValPr 3925
 QY 1195 AGAACACTGCTCCAAG---TGCTGCGAGCAGACATGATCTCTGGAGGAATTTGGAGGTG 1251
 Db 3925 oSerThrCysLysGluGlyCysLeuCysGlnProGlyTyrrPheLeuAsnAsnGlyLysCy 3945
 QY 1252 CTATCACAGATTCTCCACACTGAGTTTAAATCT-----AAGGG 1290
 Db 3945 sValLeuGln---ThrHisCysAspCysLysAspAlaGluGlyGlyLeuValProAlaGl 3964
 QY 1291 ATTGAATAATGGAGTAGAGTATAAAGTGAATGCAATGTCATGTTGATTTGCTTAGTAGAAA 1350
 Db 3964 yLys-ThrTrpThrSerLysAspCysThrGlnSerCys---AlaCysThrGlyGlyAlaV 3983
 QY 1351 TCTCTAGTTTGAAGAGGATGTTTAGGGAAACATGAGGCTGCTGCGACCAACACACAGG 1410
 Db 3983 aLgln-----CysGlnAsnPheGlnC 3990
 QY 1411 CTCCCTCGATCCCTGGGCCCCAGGAGTTTACTCAGAGCTCTCTGAAGATTTGGCAACCC 1470
 Db 3990 ysProLeu-----GlyThrTyrcysLysAspSerGlyAspGlySerA 4005
 QY 1471 ATGCCCTTTTCTGAGGAGTGCATGGCTGCTGAGCATGTTTCTGCTGCGCCAGAGAGAG 1530
 Db 4005 snCys-----ThrLysIleHis-----LysGlyA 4013
 QY 1531 AGCTTGGGTTCCTCATAGTCTGGGAGAGTGTCTGCGAGGCGCGGAGGAGCAGACA-----1586
 Db 4013 laMetGly-----AspGlyValLeuMetAlaGlyGlyIleArgAlaLeuG 4028

Db 1459 ThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuPro 1478
|||
Qy 797 AAGCATGCTGTGTAGAAATCCAGAGCTTCCAGCCAAATATGGAGGAGGCCAATTG 856
|||
Db 1479 GlnAspAlaValSerArgThrGlnArg-----ArgGlyArgThrGlyArgGlyLysPro 1496
|||
Qy 857 GGC-----AATGTCAAGCCCGACAGCAACAAAGGGTCAACTCGAGGAGTCAAGC 907
|||
Db 1497 GlyIleThrArgPheValAlaProGlyGlu-----ArgProSerGly 1510
|||
Qy 908 TGCCTGAAGAAT-----TACTGGAGTGTATGAGGCCCAAAATATGTCTTCTATT 961
|||
Db 1511 MetPheAspSerValLeuGluCysTyrAspAla----- 1523
|||
Qy 962 TGCAAATGCATGTGTTCCAAATATGAAAGAGCCAGACAGCAAGACACATTAATGAGC 1021
|||
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro-----AlaGluThrThrValArg 1538
|||
Qy 1022 ATGCCAACTACATGCAGACTGGAGTTTG----- 1051
|||
Db 1539 LeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrp 1558
|||
Qy 1052 GAAGC-----AGCATTTACCTGCCACCAAGCAATTT 1084
|||
Db 1559 GluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGln 1578
|||
Qy 1085 TCAGGA-----CTTCCAAGATTCAGTC-----ACGATGGCGGCTCTCTCA 1126
|||
Db 1579 SerGlyGluAsnLeuProTyrLeu-ValAlaTyrGlnAlaThrValCysAlaArgAlaG1 1598
|||
Qy 1127 TGCATCTCTCTGGAGGTGG-----TGGAGGCCACATCGCGCTGCCCTGTGCTCAG 1177
|||
Db 1598 nAlaProProSerTrpAspGlnMetTrpLys-----CysLeuIleArg 1613
|||
Qy 1178 GGAGAGAGCCGGAAGAACACTGCTCCAGTGCCTGCAGAGCAGATGATCCTGGAG 1337
|||
Db 1613 gLeuLysProThrLeuHisGlyProThrProLeuLeuTyrArgLeu----- 1628
|||
Qy 1238 GAATTTGGAAGGTCTTATCACAGATTCCTCCACCTCAGT----- 1277
|||
Db 1629 -----GlyAlaValGlnAsnGluIleThrLeuThrHisProValThrLysTyrI1 1645
|||
Qy 1278 -----TTAATCTAAGGATTTGAAATAGTAGTAGATATAAG 1315
|||
Db 1645 eMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyG1 1665
|||
Qy 1316 TGTGAATGCA-----TGTTGATTTCTCTAGTCTAGAAATCTCTAGTTTAAAGAGAT 1369
|||
Db 1665 yValLeuAlaAlaLeuAlaTyrCysLeuSerThrGlyCysValValIleVal----- 1683
|||
Qy 1370 GTTTAGGGAACATGAGGCTGGCTCTGCAGCAACACAGGCTCCCTGCATCCCTGGGC 1429
|||
Db 1684 -----GlyArgValValLeuSerGlyLysProAlaIleIleProAspArg 1698
|||
Qy 1430 CCAGGAGTTTACTAGAGCTCTCT-----GAAGATGTGGCAACCATGCCCTTTTC 1483
|||
Db 1698 gGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrI1 1718
|||
Qy 1484 TGAGGAGTGCATGCGCTGAGCATGTTTGTCTGGCCAGAGGAGAGAGCTTGGTTCCC 1543
|||
Db 1718 eGluGln-----GlyMetMetLeuAlaGluInPheLysGlnLysAlaLeuGlyLeu-- 1735
|||
Qy 1544 ATAGTCTGGAGAGTGTCTGCAGGCGGGGAGGCGGAGAGGCGGCTCGGAGAGCTC 1603
|||
Db 1736 -----LeuGlnThrAlaSerArgGlnAlaGluValIleAla---ProAlaValGlnTh 1752
|||
Qy 1604 ACTCTGCTGACTCTT 1619
|||
Db 1752 rAsnTrpGlnLysLeu 1757
|||
RESULT 11
LMA5_MOUSE

ID LMA5_MOUSE STANDARD; PRT; 3718 AA.
AC 061001; Q9JHQ6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LAMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpi R., Sasaki T.;
RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
chain."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-3718 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96081906; PubMed=7499364;
RA Miner J.H., Lewis R.M., Sanes J.R.;
RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
RN [3]
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
is thought to mediate the attachment, migration, and organization
of cells into tissues during embryonic development by interacting
with other extracellular matrix components.
CC -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
different polypeptide chains (alpha, beta, gamma), which are bound
to each other by disulfide bonds into a cross-shaped molecule
comprising one long and three short arms with globules at each
end.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ293593; CAB99255.1; -
DR EMBL; U37501; AAC53430.1; -
DR HSSP; P02468; LTLE.
DR MGD; MGI:105382; Lama5.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00052; laminin_B; 1.
DR Pfam; PF00053; laminin_EGF; 19.
DR Pfam; PF00054; laminin_G; 2.

QY 96 CACGTC-----GCGCTCTGCGAGCCCGCAGCCCGCCGCTGC 134
 Db 128 GlyThrGluArgTrpTrpGlnSerProLeuSerArgGlyLeuGluTyrAsnGluVal 147
 QY 135 AACGTG-----CACTTCTGCTCCGCTGCTACCGCGCAC 170
 Db 148 AsnValThrLeuAspLeuGlyGlnValPheHisValAlaTyrValLeuIleLysPheAla 167
 QY 171 CGCAGCCCGCGGGTGTTCCTCCCTGCGCGCTGCTGCGAAGGAGCCCTCCACCC 230
 Db 168 AsnSerProArgPro-----AspLeuTrpValLeuGluArgSerThrAspPhe 183
 QY 231 GGGCGTCCCGATGATCCCATGTAAGTAAAGTAAGCAGGTGTACTACAAAGTAATA 290
 Db 184 GlyHisThrTrpGlnPro-----TrpGlnPhePhe 193
 QY 291 ATCCGGAAGAGCAACTTTCAGNATCTTCTGCTCAGGAATCCTGTTGCAAGTTCCCAT 350
 Db 194 AlaSerSerLysArgAspCys-----LeuGluArgPhe 204
 QY 351 GGTCCAGGAAGTACAGGATCCCTGCTCTTCTTAAGAAAGATTCACCAACCCATGG 410
 Db 205 GlyProArgThr-----Leu-GluArgIleThrGlnAspAs 216
 QY 411 TGAT-----ATCCCAATTGAAGGGGCGACACAAA 440
 Db 216 pAspValIleCysThrThrGluTyrSerArgIleValProLeuGluAsnGlyGluIle-V 236
 QY 441 TGCTATGTATAGCAATCTAGACACAGAGACTAAAGCAGCTCCATTTGGTTCCTCAGT 500
 Db 236 alValSerLeuValAsnGlyArgProGlyAlaLeu----- 247
 QY 501 ATCAAGATCAAAATAATTATCTACAGTCAGATGTCCTTAAACCAATGACTGCTTTAGTAG 560
 Db 248 -----AsnPheSerTyrSerProLeuLeuArgAspPheThrLysAlaThrA 263
 QY 561 GG-----AGATTTTTCGAGCATCAACAAATTA-----AATCTCATACACAAAC 608
 Db 263 snIleArgLeuArgPheLeuArgThrAsnThrLeuLeuGlyHisLeuMetGlyLysAlaL 283
 QY 609 TTGAGGAGCCTTACCATCGGTAGTCAACGGTCTGCTTCCCTCGGGATCACTCTTC 668
 Db 283 euArgAsp-----ProThrValThrArgArgTyrTyr----- 294
 QY 669 CAGGACCACCAAAATAACTTTGGCTGGTACTGTGACTGCTTTGCCAGTGGGACTTTT 728
 Db 295 --SerIleLysAspIleSerIleGlyArgCysValCysHisGlyHisAlaAspValC 314
 QY 729 GC-----AACAACT 737
 Db 314 ysAspAlaLysAspProLeuAspProPheArgLeuGlnCysAlaCysGlnHisAsnThrC 334
 QY 738 GC-----AATTGTAATAATTGTGCAACACTTGCATCATGATATATGAAGGTTTAAAG 791
 Db 334 ysGlyGlySerCysAspArgCysCysProGlyPheAsnGln-----GlnProTrpLysP 352
 QY 792 CCATTAAGGCA-----TGTCTGGT----- 811
 Db 352 roAlaThrThrAspSerAlaAsnGluCysGlnSerCysAsnCysHisGlyHisAlaTyrA 372
 QY 812 -----AGAAATCCAGAGCTTTCACGCCCAAAATTTG 842
 Db 372 spCysTyrTrpAspProGluValAspArgAsnAlaSerGlnAsnGlnAspAsnValT 392
 QY 843 GGAAGGCCAATTGGCAATGTC-----AAGCCCCAGCACACAAAGGGTGCACT 893
 Db 392 yrGlnGly-----GlyGlyValCysLeuAspCysGlnHisThrGlyIleAsnC 410
 QY 894 CGAGGAGGTGAGGTGCTCAAGAAATTAATGCGAGTCTATGAGGCCCAATATATGCTT 953
 Db 410 ysGluArg-----CysLeuProGlyPhePheArgAlaProAspGlnProLeuAspSerP 428
 QY 954 CTTCTATTGCAATGCAATGGTTGCAAAAT-----TATGAAGAAA 995

Db 428 roHisValCysArgProCysAspCysGluSerAspPheThrAspGlyThrCysGluAspL 448
 QY 996 GCCCAGAACCAAGACACTAATGAGCATGCCAAACTACATGCAGACTGGAGGTTTG--- 1051
 Db 448 euThrGlyArg-----CysTyrCysArgProAsnPhe-----ThrGlyGluLeuCysA 464
 QY 1052 -----GAAGGC-----AGCCATTACCTGCCACCAACAAATTTTCAG 1088
 Db 464 laAlaCysAlaGluGlyTyrThrAspPheProHisCysTyrPro----- 478
 QY 1089 GACTTCCAGATTTCAGTCAC---CATAGGCGGCTTCTCTCATGCTCTCTGGAGGTGG 1145
 Db 479 --LeuProSerPheProHisAsnAspThrArgGluGlnValLeuProAlaGlyGlnIleV 498
 QY 1146 TGGAGGCCACATCGCTGCTGCTCAGGGAAGAGCGCGGAGAAAGAA--- 1198
 Db 498 alaAsnCysAspCysAsnAlaAlaGlyThrGlnGlyAsnAlaCysArgLysAspProArgL 518
 QY 1199 -----CACTGCTCCAACTGC----- 1213
 Db 518 euGlyArgCysValCysLysProAsnPheArgGlyAlaHisCysGluLeuCysAlaProG 538
 QY 1214 -----CTGGCAGACGAGA 1226
 Db 538 lyPheHisGlyProSerCysHisProCysGlnCysSerSerProGlyValAlaAsnSerL 558
 QY 1227 TGATCTGGAGGAATTTGGAAGGTGCTTA 1255
 Db 558 euCysAspProGluSerGlyGlnCysMet 567
 RESULT 12
 ID NTC4_HUMAN STANDARD: PRT: 2003 AA.
 AC Q99466; Q00306; Q99940; Q99458; Q9H3S8; Q9U1I9; Q9U1J0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 DE (hNotch4).
 GN NOTCH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
 RC TISSUE-Placenta;
 RX MEDLINE-97311416; PubMed-9168133;
 RA Sugaya K., Sasanuma S.-I., Nohata J., Kimura T., Fukagawa T.,
 RA Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
 RT "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
 human counterpart gene of mouse proto-oncogene Int3";
 RL Gene 189:235-244(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE-Bone marrow, and Heart;
 RX MEDLINE-98360091; PubMed-9693032;
 RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
 RA Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 sequence of the human NOTCH4 gene";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
 RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE-99180765; PubMed-10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;

RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -|- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -|- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -|- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; may be
 CC produced by alternative splicing.
 CC -|- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -|- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -|- PTM: Phosphorylated (By similarity).
 CC -|- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -|- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -|- SIMILARITY: CONTAINS 28 EGF-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -|- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -|- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D63395; BAA09708.1; ALT_FRAME.
 CC EMBL; D86566; BAA13116.1; -.
 CC EMBL; U95299; AAC32288.1; -.
 CC EMBL; U89335; AAC63097.1; -.
 CC EMBL; AB023961; BAB20317.1; -.
 CC EMBL; AB024520; BAA88951.1; -.
 CC EMBL; AB024578; BAA88952.1; -.
 CC HSP; P08709; IBF9.
 CC Genew; HGNC:7884; NOTCH4.
 CC MIM; 164951; -.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR000742; EGF-2.
 CC InterPro; IPR001881; EGF_CA.
 CC InterPro; IPR001438; EGF-II.
 CC InterPro; IPR000800; Notch.
 CC Pfam; PF00008; EGF; 26.
 CC Pfam; PF00023; ank; 6.
 CC Pfam; PF00066; notch; 2.
 CC PRINTS; PR00010; EGFBLD00.
 CC
 CC PRINTS; PR00011; EGFELAMIN.
 CC PRINTS; PR00012; FNTYPEI.
 CC SMART; SM00248; ANK; 5.
 CC SMART; SM00179; EGF_CA; 11.
 CC SMART; SM00001; EGF_Like; 15.
 CC SMART; SM00004; NL; 2.
 CC PROSITE; PS50088; ANK_REPEAT; 5.
 CC PROSITE; PS50297; ANK_REPEAT; 5.
 CC PROSITE; PS00010; ASX_HYDROXYL; 11.
 CC PROSITE; PS00022; EGF_1; 28.
 CC PROSITE; PS01186; EGF_2; 21.
 CC PROSITE; PS01187; EGF_CA; 9.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 CC Triplet repeat expansion; Alternative splicing.
 CC SIGNAL 1 23 POTENTIAL.
 CC CHAIN 24 2003 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
 CC CHAIN 1432 2003 NOTCH EXTRACELLULAR TRUNCATION (BY
 CC SIMILARITY).
 CC CHAIN 1467 2003 NOTCH INTRACELLULAR DOMAIN (BY
 CC SIMILARITY).
 CC DOMAIN 24 1447 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 1448 1468 POTENTIAL.
 CC DOMAIN 1469 2003 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 24 63 EGF-LIKE 1.
 CC DOMAIN 64 115 EGF-LIKE 2.
 CC DOMAIN 118 135 EGF-LIKE 3.
 CC DOMAIN 156 192 EGF-LIKE 4.
 CC DOMAIN 194 232 EGF-LIKE 5.
 CC DOMAIN 234 274 EGF-LIKE 6.
 CC DOMAIN 276 312 EGF-LIKE 7.
 CC DOMAIN 314 353 EGF-LIKE 8.
 CC DOMAIN 355 391 EGF-LIKE 9.
 CC DOMAIN 392 430 EGF-LIKE 10.
 CC DOMAIN 432 473 EGF-LIKE 11.
 CC DOMAIN 475 511 EGF-LIKE 12.
 CC DOMAIN 513 549 EGF-LIKE 13.
 CC DOMAIN 551 587 EGF-LIKE 14.
 CC DOMAIN 589 625 EGF-LIKE 15.
 CC DOMAIN 626 659 EGF-LIKE 16.
 CC DOMAIN 661 689 EGF-LIKE 17.
 CC DOMAIN 691 727 EGF-LIKE 18.
 CC DOMAIN 729 765 EGF-LIKE 19.
 CC DOMAIN 767 803 EGF-LIKE 20.
 CC DOMAIN 806 842 EGF-LIKE 21.
 CC DOMAIN 844 880 EGF-LIKE 22.
 CC DOMAIN 882 928 EGF-LIKE 23.
 CC DOMAIN 930 966 EGF-LIKE 24.
 CC DOMAIN 968 1004 EGF-LIKE 25.
 CC DOMAIN 1006 1044 EGF-LIKE 26.
 CC DOMAIN 1046 1085 EGF-LIKE 27.
 CC DOMAIN 1087 1126 EGF-LIKE 28.
 CC DOMAIN 1130 1171 EGF-LIKE 29.
 CC DOMAIN 1172 1212 POLY-ARG.
 CC REPEAT 1165 1212 LIN/NOTCH 1.
 CC REPEAT 1213 1246 LIN/NOTCH 2.
 CC REPEAT 1247 1286 LIN/NOTCH 3.
 CC REPEAT 1633 1665 ANK 1.
 CC REPEAT 1666 1698 ANK 2.
 CC REPEAT 1700 1732 ANK 3.
 CC REPEAT 1733 1765 ANK 4.
 CC REPEAT 1766 1798 ANK 5.
 CC DISULFID 28 41 BY SIMILARITY.
 CC DISULFID 35 51 BY SIMILARITY.
 CC DISULFID 53 62 BY SIMILARITY.
 CC DISULFID 68 80 BY SIMILARITY.
 CC DISULFID 74 103 BY SIMILARITY.
 CC DISULFID 105 114 BY SIMILARITY.
 CC DISULFID 122 133 BY SIMILARITY.
 CC DISULFID 127 143 BY SIMILARITY.
 CC DISULFID 145 154 BY SIMILARITY.
 CC DISULFID 160 171 BY SIMILARITY.

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate BK) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11105;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91140698; PubMed=1847440;
RA Takemizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
RA Onishi E., Andoh T., Yoshida I., Okayama H.;
RT "Structure and organization of the hepatitis C virus genome isolated
RT from human carriers.";
RL J. Virol. 65:1105-1113(1991).
RN [2]
RP SEQUENCE OF 1487-1500.
RX MEDLINE=96235224; PubMed=8647104;
RA Borowski P., Heiland M., Oehmann K., Becker B., Kornetky L.;
RT "Non-structural protein 3 of hepatitis C virus inhibits
RT phosphorylation mediated by cAMP-dependent protein kinase.";
RL Eur. J. Biochem. 237:611-618(1996).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
RX MEDLINE=97015088; PubMed=8861916;
RA Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,
RA Moonaw E.W., Adachi T., Hostomsky J.;
RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
RT trypsin-like fold and a structural zinc binding site.";
RL Cell 87:331-342(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
RX MEDLINE=98227846; PubMed=9568891;
RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
RA Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;
RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
RL Protein Sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M58335; AA472945.1; -
DR PIR: A38465; GNWVTC.
DR PDB: 1AIQ; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR PDB: 1NS3; 08-APR-98.
DR MEROPS: S29.001; -
DR MEROPS: U39.001; -
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4A.
DR InterPro: IPR001490; HCV_NS4B.
DR InterPro: IPR002868; HCV_NS5A.
DR InterPro: IPR002166; HCV_RDRP.

DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR Pfam: PF00998; HCV_RDRP; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR ProDom: PD186062; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 131 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 132 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1852 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 136 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327189 MW; P8422D5ECCFDF9C CRC64;

Alignment Scores:

Pred. No.:	0.386	Length:	3010
Score:	116.00	Matches:	93
Percent Similarity:	34.88%	Conservative:	57
Best Local Similarity:	21.63%	Mismatches:	157
Query Match:	3.03%	Indels:	123
DB:	1	Gaps:	23

US-09-743-237-3 (1-2134) x POLG_HCVBK (1-3010)

QY	542	CCAATGACGCTTTAGTAGGAGA-----	TTTTTCCAGCATCAACAAATTA	589
		: :		
Db	1381	ProileglualleargGlyArgHisLeullepheCysHisserLysLysCys		1400
QY	590	AATCTCATTACACAACAACCTTGAGGAGCCCTTACCATCGGTGACACGGGTCTGCTTC		649
		: :	: :	: :

Db 1401 AspGluLeuAlaAlaLysLeuSerGlyLeu-----GlyIleAsnAlaValAlaTyr 1417
QY 650 CCCTCGGGATCAACTCTTCAGGACCCACCAAAATA----- 685
Db 1418 TyrArgGlyLeuAspValSerValIleProThrIleGlyAspValValValAlaThr 1437
QY 686 -----ACTTTGGCTGGGTACTGTGACTGCTTGCAGTGGGACTTTTGCACAACTGC 739
Db 1438 AspAlaLeuMetThrGlyTyr-----ThrGlyAspPheAspSerValIle 1452
QY 740 AATGTAAATAATTTCTCAACAACCTTG-----CATCATGATATT 778
Db 1453 AspCysAsnThrCysValThrGlnThrValAlaAspPheSerLeuAspProThrPheThrIle 1472
QY 779 GAACGGTTTAAAGCCATTAAGCATGCTTGGTAGAATCCAGAAAGCTTTCAGACCAAAA 838
Db 1473 GluThrThrThrValProGlnAspAlaValSerArgSerGlnArg-----ArgGlyArg 1490
QY 839 ATTGGGAAGGCCAATTTGGGC-----AATGTCAGAGCCCGCAGCACAACAAGGGTGC 889
Db 1491 ThrGlyArgGlyArgArgGlyIleTyrArgPheValThrProGlyGlu----- 1506
QY 890 AACTGCAGGAGCTCAGCTCCCTCAAGAAAT-----TACTGCCAGTCTATGAGGCCCAA 943
Db 1507 -----ArgProSerGlyMetPheAspSerValLeuCysGluCysTyrAspAla--- 1523
QY 944 ATTATGTGTTCTTCTATTTCAAATGATTTGCTTGCAGAAATATGAAGAAAGCCCAAGAA 1003
Db 1524 -----GlyCysAlaTrpTyrGluLeuThrPro--- 1532
QY 1004 CGAAGACACTAATGACATGCCAACTACATGACGAGGTGGAA----- 1054
Db 1533 AlaGluThrSerValArgLeuAlaTyrAlaTyrLeuAsnThrProGlyLeuProValCysGln 1552
QY 1055 -----GCCAGCCATTAC 1066
Db 1553 AspHisLeuGluPheTrpGluSerValPheThrGlyLeuThrHisIleAspAlaHisPhe 1572
QY 1067 CTGCCACCAACGAAATTTTCAGGA-----CTTCCAGATTCAGTC-----AC 1108
Db 1573 LeuSerGlnThrLysGlnAlaGlyAspAsnPheProTyrLeu-ValAlaTyrGlnAlaThr 1592
QY 1109 GATAGGCGGCTTCTCATGCTCTCTGGAGGTGG-----TGGAGGCCACATGC 1159
Db 1592 rValCysAlaArgAlaGlnAlaProProSerTrpAspGlnMetTrpLys----- 1609
QY 1160 GCCTGCTGCTGCTCAGGAGGAGGCGCGAGAAAGAACTGCTCCAAGTGCCTGCGCA 1219
Db 1610 -----CysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeuTyrAr 1627
QY 1220 GAGCAGATGATCCTGGAGGAATTTGGAAGGTGCTTATCACAGATTCCTCCACACTGAGT-- 1277
Db 1627 gLeu-----GlyAlaValGlnAsnGluValThrLeuThrHi 1639
QY 1278 -----TTAATCTAAGGATTCGAAA 1297
Db 1639 sProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrSerTh 1659
QY 1298 ATGGAGTAGAGTAAAGTCTGAATGATGCTGATTTTGTCTAGTAGAATCTCTAG 1357
Db 1659 rTrpValLeuValGlyValLeuAlaLeuAlaAlaTyrCysLeuThrThrGlyse 1679
QY 1358 TTTAGAAGGATGTTTGGGGAAACATGAGGCTGGCTTCGACGACCAACACAGGCTCCCTC 1417
Db 1679 rValValIleVal-----GlyArgIleIleLeuSerGlyArgProAlaIle 1694
QY 1418 GCATCCCTGGCCCGGAGGAGTTTACTCAGAGCTCTCT-----GAAGATGTGGCAACCCA 1471
Db 1694 eValProAspArgGluLeuLeuTyrGlnGluPheAspGluMetGluGluCysAlaSerHi 1714
QY 1472 TGCCCCCTTTCTCAGGAGGTGCATGGCTGAGCATGTTTGTCTGCCCGCAGGAGGAGA 1531
Db 1714 sLeuProTyrIleGluGln-----GlyMetGlnLeuAlaGluGlnPheLysGlnLysAl 1732

QY 1532 GCTTGGGTTCCCATAGTCTCGGAGAGTGTCTGCAGGGCGCGGAGGCGAGAGCAGGCC 1591
Db 1732 aLeuGlyLeu-----LeuGlnThrAlaThrLysGlnAlaGluAla---AlaAlaPr 1748
QY 1592 TCGCGAGAGCTCACTCTGGTCGACTCTT 1619
Db 1748 oValValGluSerLysTrpArgAlaLeu 1757
RESULT 14
POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals".
RL Virus Res. 23:39-53(1992).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC {RNA}(N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D11168; BAA01943.1;
CC PIR: A45573; A45573.
CC HSSP: P26663; IJXP.
CC MEROPS: S29.001; -.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRP.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; Y00057; CAA68266.1; -
DR EMBL; X64072; CAA45427.1; -
DR EMBL; X64073; CAA45427.1; JOINED.
DR EMBL; X64074; CAA45427.1; JOINED.
DR EMBL; X64075; CAA45427.1; JOINED.
DR EMBL; X64076; CAA45427.1; JOINED.
DR EMBL; X64077; CAA45427.1; JOINED.
DR EMBL; X64078; CAA45427.1; JOINED.
DR EMBL; X64079; CAA45427.1; JOINED.
DR EMBL; X64080; CAA45427.1; JOINED.
DR EMBL; X64081; CAA45427.1; JOINED.
DR EMBL; X64082; CAA45427.1; JOINED.
DR EMBL; X64083; CAA45427.1; JOINED.
DR EMBL; X63924; CAA45427.1; JOINED.
DR EMBL; X63925; CAA45427.1; JOINED.
DR EMBL; X63926; CAA45427.1; JOINED.
DR EMBL; AL163300; CAB90553.1; -
DR EMBL; M15395; AAS59490.1; -
DR EMBL; S81234; AAB21404.1; -
DR PIR; A25967; IJHULM.
DR PIR; A29265; A29265.
DR HSSP; P05106; IJV2.
DR Genew; HGNC:6155; ITGB2.
DR MTM; 600065; -
DR MIM; 116920; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002369; Integrin_B.
DR InterPro; IPR001169; Integrin_beta_C.
DR InterPro; IPR003659; Plexin-like.
DR InterPro; IPR002035; WWF-A.
DR Pfam; PF00362; Integrin_B.
DR PRINTS; PR0186; INTEGRINB.
DR ProDom; PD00181; Integrin_B; 1.
DR SMART; SM00187; INB; 1.
DR SMART; SM00423; PST; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00243; INTEGRIN_BETA; 3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
Repeat; Signal; Disease mutation.
FT SIGNAL 1 22
FT CHAIN 23 769
FT DOMAIN 23 700
FT TRANSMEM 701 723
FT DOMAIN 724 769
FT DOMAIN 124 363
FT DOMAIN 449 617
FT REPEAT 449 496
FT REPEAT 497 540
FT REPEAT 541 581
FT REPEAT 582 617
FT DISULFID 25 447
FT DISULFID 33 43
FT DISULFID 36 73
FT DISULFID 46 62
FT DISULFID 191 198

Alignment Scores:
Pred. No.: 0.32 Length: 769
Score: 115.50 Matches: 59
Percent Similarity: 34.16% Conservatives: 37
Best Local Similarity: 21.00% Mismatches: 101
Query Match: 3.02% Indels: 84
DB: 1 Gaps: 16

US-09-743-237-3 (1-2134) x ITB2_HUMAN (1-769)

QY 304 AACTTTCAGAAATCTTCTGCTCAGGAATCCTGTTCAGAGTCTCCATGTGCTCCAGGAACT 363
Db 394 AsnGlnProArgGlyAspCysAspGlyValGlnIleAsnValProIle-ThrPheGlnVa 413
QY 364 AGAGGATCCCTCTGCTGTTCTTAAAGAAGATTCC----- 400
Db 413 LysValThrAlaThrGluCysIleGlnGluInSerPheValIleArgAlaLeuGlyPh 433
QY 401 -AACCCCAATGGTGATATGCAATTTGAAAGGGGGCACACAATGCTATGTATAGCAATTC 459
Db 433 eThrAspIleValThrValGlnValLeuProGlnCysGluCysArgCysArgaspGlnSe 453
QY 460 TAGACAAGAGAACTAAAGCAGCTCCATTGGTTCTCTAGTATCAAGATCAAAATAATTA 519
Db 453 rArgAspArgSerLeu-----CysHis-----GlyLysGlyPh 464
QY 520 TCTACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGGAGAGATTTTGGCAGCATC 579
Db 464 eLeuGluCysGlyIleCysArgCysAspThrGlyTyrIleGlyLysAsnCysGluCysG1 484
QY 580 AACAAAATTAATCTCATTACACAACAACCTTGAGGGAGCC----- 619
Db 484 nThrGln---GlyArgSerSerGlnGluLeuGluGlySerCysArgLysAspAsnAsnSe 503
QY 620 -----TTACCATCGGTAGTCAACGGGCTCTGCTTTCCCTCGGGATCAAC 663
Db 503 rIleIleCysSerGlyLeuGlyAspCysValCysGlyGlnCysLeuCysHisThrSerAs 523
QY 664 TCTTCAGAGACCACCAAAATAACTTTGGCTGGGTACTGTGACTGTGCTTTGCC----- 715
Db 523 pValProGly-----LysLeuIleTyrGlyGlnTyrCysGluCysAspThrIleAsnCy 541
QY 716 -----AGTGGGAGCTTTTCCACAAC-----TGCAATTGTAATAA 750
Db 541 sGluArgTyrAsnGlyGlnValCysGlyGlyProGlyArgGlyLeuCysPheCysGlyLy 561
QY 751 TTGTTGCACAACACTTGTCATCATGATATTGAACGGTTT-----AAAGCCAT 795
Db 561 sCys-----ArgCysHisProGlyPheGluGlySerAlaCysGlnCysGluArgThrTh 579
QY 796 TAAGGATGCTTTGGTAGAAATCCAGAAAGCTTTCCAGCCCAAAATTTGGGAAGGCCAATT 855
Db 579 rGluGlyCysLeu-----AsnProArgArgValGluCysSer---GlyArgGlyArg-- 595
QY 856 GGCAATGTCAAGCCCCAGCACACAAGGGTGCACACTGCAGGAGGTGCAGCTGCCTGAA 915
Db 596 -----CysArgCys----- 598
QY 916 GAATTACTGCGAGTGTATGAGGCCCAAAATTATG-----TG 951
Db 599 -AsnValCysGluCysHisSerGlyTyrGlnLeuProLeuCysGlnGluCysProGlyCy 618
QY 952 TTCTTCTATTTC-----AAATGCATTGGTTGCAAAATTTATGAAGAAAGCC 999
Db 618 sProSerProCysGlyLysTyrIleSerCysAlaGluCysLeuLysPheGluLysGlyPr 638
QY 1000 A 1000
Db 638 o 638

Search completed: April 21, 2003, 11:57:01
Job time : 90.5 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 21, 2003, 11:50:49 ; Search time 136.5 Seconds

(without alignments)
6442.557 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 3824

Sequence: 1 aattcgggtcgaaggcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=rlp
-Q/cgn2_1/USTPO.spool/US0974337/runat_21042003_112209_1473/app-query.fasta_1.2311
-DB=SPTREMBL_21 -Qfmt=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0974337.cgn_1.1.238 -runat_21042003_112209_1473 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Databse :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1515.5	39.6	475	11 Q8VIE0	Q8vie0 mus musculus

ID	Q8VIE0	PRELIMINARY;	PRT;	475 AA.
AC	Q8VIE0	Q8VIE0	Q8VIE0	Q8VIE0
DT	01-MAR-2002	(TRENBLrel. 20, Created)		
DT	01-MAR-2002	(TRENBLrel. 20, Last sequence update)		
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)		
DE	Tesmin	(Hypothetical 50.6 kDa protein).		
GN	TESM			
OS	Mus musculus	(Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALE/C;			
RA	Suton S., Mwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;			
RT	"Structure of the tesmin gene encoding a testis-specific persistent protein; a possible multifunctional protein with dynamic changes of			
RT	localization throughout spermatogenesis."			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			

ALIGNMENTS

RESULT 1

ID	Q8VIE0	PRELIMINARY;	PRT;	475 AA.
AC	Q8VIE0	Q8VIE0	Q8VIE0	Q8VIE0
DT	01-MAR-2002	(TRENBLrel. 20, Created)		
DT	01-MAR-2002	(TRENBLrel. 20, Last sequence update)		
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)		
DE	Tesmin	(Hypothetical 50.6 kDa protein).		
GN	TESM			
OS	Mus musculus	(Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BALE/C;			
RA	Suton S., Mwa K., Matsuura T., Kawasaki Y., Ohinata Y., Mitsui Y.;			
RT	"Structure of the tesmin gene encoding a testis-specific persistent protein; a possible multifunctional protein with dynamic changes of			
RT	localization throughout spermatogenesis."			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.			


```

Db 162 erSerGlnGluAlaGluGluAlaSerCysProArgLysLysAspSerProMetv 182
QY 411 TGATATGCCAATTAAGAGGGGGCACCAAAATGCTATGATAGACAATTTAGAACAGAG 470
Db 182 allleCysGlnLeuLysGlyGlyAlaGlnMetLeuCysLeuAspCysGlyAlaArg 202
QY 471 AACTAAAGCACTCCATTTGGTCCCTCAGTATCAAGATCAAAATAATATCTACAGTCAG 530
Db 202 luLeuLysAlaLeuHisLeuLeuProGlnTyrAspAspGlnSerPheProGlnSerG 222
QY 531 ATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTGGCCAGCATCAACAAATTA 590
Db 222 luLeuProLysProMetThrThrLeuValGlyArgLeuLeuProValProAlaLysLeu 242
QY 591 ATCTCATTTACACAACTTGAGGGAGCCCTTACCTCGGTAGTCAACGGGTCTGCTTTCC 650
Db 242 snLeuLeuThrGlnValAspAsnGlyAlaLeuProSerAlaValAsnGlyAlaAlaPhe 262
QY 651 CTCTGGGATCACTCTCCAGGACCACCACCAAAATAATTTGGCTGGG 697
Db 262 roSerGlyProAlaLeuGlnGlyProProLysIleThrLeuSerGly 277

RESULT 4
Q9V6Q8
ID Q9V6Q8 PRELIMINARY; PRT; 950 AA.
AC Q9V6Q8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG6061 protein.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003818; AAF58365.1;
DR FlyBase; FBgn0033846; CG6061.
DR InterPro; IPR005172; CXG.
DR Pfam; PF03638; CXG.
SQ SEQUENCE 950 AA; 100021 MW; 627C3EA6B4A0A30 CRC64;

Alignment Scores:
Pred. No.: 1,01e-38 Length: 950
Score: 483.50 Matches: 155
Percent Similarity: 39.73% Conservative: 81
Best Local Similarity: 26.09% Mismatches: 168
Query Match: 12.64% Indels: 190
DB: 5 Gaps: 18

US-09-743-237-3 (1-2134) x Q9V6Q8 (1-950)
QY 1 AATTCGGGGTCAAGCGCAAGCTCGCGGGGGCGACAGCGCGGGG-----AGC 51
Db 357 AsnLysAsnIleLeuLysLeuSerProAsnAlaSerAlaThrSerSerThrHisThr 376
QY 52 TCCTCGGGGAGTACCCCGGGATCCAGAGC----- 81
Db 377 ThrGlyGlnThrProSerSerGlyLeuHisAlaIleGlnLeuProGlyLys 396
QY 82 -----TCAGCGCGCTGGAGGAGCTGC 102
Db 397 GlyGlyIleGlnTyrValArgValLeuAsnAsnAsnLysSerAlaAlaGlyThrSerAla 416
QY 103 CCTCTCTGAGCGCCCGCAGC-----CGCCCGCTGCA 135
Db 417 ThrAlaSerIleProLysThrValGlnThrGlnLysIleThrValValArgProAla 436
QY 136 -----ACGTGCACCTTCCTGCTCGCTGCTACCGCGCGCACC 171
Db 437 AlaThrGlyValProAlaThrSerThrThrThrSerAlaAlaAlaSerProAlaAla 456
QY 172 GCAGCGCGGGGTGTTTCCCGCTGG----- 198
Db 457 AlaSerLysAlaAsnLeuAlaMetGlyAsnThrAsnLysIleValMetArgSerMetGly 476
QY 199 -----CGCGCTGGGTCC 210
Db 477 GlySerIleValProLeuProSerValGlnThrLeuValSerLysArgAlaLeuGlyAla 496
QY 211 TCCGAAGGAGCGCTCCACCGCGCGCTCCGATGATCC----- 247
Db 497 IleSerAsnAlaSerLysPro-AlaSerAlaAlaSerSerSerSerAlaThrProSerAlaSe 516
QY 248 -----CAGTTGAATTCAGGTA 264
Db 516 rGlnGluLeuProArgLysHisArgLeuThrAspLeuAsnValGlnLeuLysGlnSerAl 536
QY 265 AGCAGGTGTACTACTACAAAGTAATAAT-----CCGGAAGAAGCAACTTTGCA 312
Db 536 aserValSerSerGluAlaSerAspSerSerAspAlaGlyProGluAlaLysLysProAr 556
QY 313 GAATCTTCTGCTCAGGAATCTGTTGCAAGTTCCTCCAGGAACTAGAG----- 367
Db 556 gTyrValIleThrMetGlnGlnGlySerGlnLysAlaAlaSerGlnProValGlnLysLe 576
QY 368 -----GATGCTCTCTGCTG 381
Db 576 uIleAsnArgThrAlaAsnValGlnArgValValSerSerSerSerThrSerProSerSer 596
QY 382 TTCTCTTAAGAAAGATTCCCAACCCCAATGATGATATATGCAATTAAGAAAGGGGCGACCAAT 441
Db 596 nSerThrLysLysIleTyrAsnTyrValGlnProThrGlySerAsnGlyAlaLysTyrMe 616

```


OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Town C.D.,
 Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 Cronin L.A., Shen M., Vankken S.E., Umayam L.J., Gill J.E.,
 Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 Salzberg S.L., Fraser C.M., Venter J.C.;
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana";
 RT Nature 402:761-768(1999).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC006081; AAD24386.1; -
 DR InterPro; IPR005172; CXC.
 DR Pfam; PF03638; CXC; 2.
 SQ SEQUENCE 571 AA; 61312 MW; 36701C5FA62B3B89 CRC64;

Alignment Scores:
 Pred. No.: 4.06e-28 Length: 571
 Score: 376.00 Matches: 95
 Percent Similarity: 44.27% Conservative: 48
 Best Local Similarity: 29.41% Mismatches: 90
 Query Match: 9.83% Indels: 90
 DB: 10 Gaps: 10

US-09-743-237-3 (1-2134) x Q9SL70 (1-571)

QY 527 TCAGATGCTCCCTAAACCAATGACTCTTAGTAGGAGATTTTCCAGCATCAACAAA 586
 Db 51 AlaserileProSerProileValThrValThrArgProIleileThrSerGlnAlaPro 70
 QY 587 TTAATCTCATTAACA-----CAACAACCTTGAGGAGCCCTTA----- 622
 Db 71 ProThrValAlaThrProileProProGlnSerGlnGlyIleileuHisVal 90
 QY 623 -----CCATCGGTACTCAACGGGTCTGCTTCCCTCCGCGATCACT 664
 Db 91 ProfileArgHisProArgProGluSerProAsnSerMetProArgProAlaGlyGluThr 110
 QY 665 CTTCCAGCAGCACCAAAA-----ATAACTTTGGCT 694
 Db 111 ArgaspGlyThrProGlnLysLysLysLysLysLysLysLysLysLysLysLys 130
 QY 695 GGGTACTGCTGCTGCTTCCAGTGGGACTTTTCAACAACCTGCAATTTGAATTAATGT 754
 Db 131 LeuTyrCysGluCysPheAlaSerGlyThrTyrCysAspGlyCysAsnLysLysLys 150
 QY 755 TGCACAACTTGCATCATGATTTGAACGGTTTAAGCCATTAAAGCCATTAGGCGATCTTGGTGA 814
 Db 151 PheAsnAsnValGluAsnGluProAlaArgArgGlnAlaValGluSerThrLeuGluArg 170
 QY 815 AATCCAGAGCTTTCCAGCCAAAATT-----GGGAAG 847
 Db 171 AsnProAsnAlaPheArgProLysLysLysLysLysLysLysLysLysLysLysLys 190
 QY 848 GCCCAATTGGCAATGTC-----AAGCCCGCAGCACAAGGCTGCAACTGCAGGAGG 901
 Db 191 GluGluValGlyAspValValMetLeuAlaArgHisAsnLysGlyCysHisCysLysLys 210
 QY 902 TCAGGCTCCCTGAACAATTTACTGGAGTGCATATGAGGCCCAAAATTTATGTTCTTCTATT 961
 Db 211 SerGlyCysLeuLysLysTyrCysGluCysPheGlnAlaAsnIleLeuCysSerGluAsn 230

QY 962 TCGAATGCATTGTTGCAAAAATTTATGAAGAAAGCCAGCAACGAAAGACACATAATAGC 1021
 Db 231 CysLysCysLeuAspCysLysAsnPheGluGlySerGluValArgGlnSerLeuPheHis 250
 QY 1022 ATGCCAAAC-----TACATGCAG-----ACTGGAGGT 1048
 Db 251 GlyGluHisSerHisAsnLeuAlaTyrLeuGlnHisAlaAsnAlaAlaIleThrGlyAla 270
 QY 1049 TTGGAGCAGCACCATTACTG-----CCACCAAGCAAAATTTTCAGGA----- 1090
 Db 271 IleGlySerSerGlyPheAlaSerAlaProProLysArgArgLysGlnGluLeu 290
 QY 1091 -----CTTCCAAGATTACGTACCGAT 1111
 Db 291 PhePheAsnGlnGlyThrLysAspSerSerThrHisArgLeuGlyGlnAlaAsnAsnGly 310
 QY 1112 AGCGGCGCTTCTCTCA-----TGCATCTCTGGGAGGTGGTGGAGGCCACA 1156
 Db 311 ArgThrThrSerSerGlnThrGlySerArgAlaGlyGlyAsnAlaSerLeuGlyProSer 330
 QY 1127 -----TGCATCTCTGGGAGGTGGTGGAGGCCACA 1156
 Db 331 LysValValTyrLysSerLeuLeuAlaAsnIleIleLysProMetAspValLysAlaLeu 350
 QY 1157 TGCAGCTGCTGCTTGTCTCAGGAGAGAGCGGAGAGAACACTGCTCCCAAGTGCCTG 1216
 Db 351 CysSerValLeuValAlaValAlaGlyGluAlaAlaLysThrLeuThrGluLysArgLeu 370
 QY 1217 GCAGAGCAG 1225
 Db 371 AlaAsnGln 373
 RESULT 8
 Q9SZD1 PRELIMINARY; PRT; 603 AA.
 ID Q9SZD1;
 AC Q9SZD1;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative transcription factor.
 OS F19B15.30 OR AT4G29000.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duisterhoef A.,
 Bancroft I., Meves H.W., Mayer K.F.X., Lemcke K., Schueller C.;
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duisterhoef A.,
 Meves H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL078470; CAB43914.1; -
 DR EMBL; AL161574; CAB79658.1; -
 DR InterPro; IPR005172; CXC.
 DR Pfam; PF03638; CXC; 2.
 SQ SEQUENCE 603 AA; 64635 MW; 7FD1951AFE818DE5 CRC64;

Alignment Scores:
 Pred. No.: 2.52e-26 Length: 603
 Score: 358.00 Matches: 93

Percent Similarity:	44.52%	Conservative:	41
Best Local Similarity:	30.90%	Mismatches:	76
Query Match:	9.36%	Indels:	91
DB:	10	Gaps:	12
US-09-743-237-3 (1-2134) x Q95SD1 (1-603)			
QY	521	CTACAGTCAGATGTCCTT-----AAACCA---ATGACTGCTTTAGTAGGAGAGATTT	568
DB	58	IleSerSerValProSerThrIleArgProGlyMetThrIleAlaIleGlyGlnVal	77
QY	569	-----TTCCAGCATCAACAAATAAATTCATATACACAACTT	610
DB	78	ThrGlnValArgProThrLeuProMetAlaThrThrMetSerAsnProProSerGlnSer	97
QY	611	GAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTCCCTCGGATCAACTCTTCCA	670
DB	98	Gln-----IleValAsnAla-----ProIleArgHisProIlePro	109
QY	671	GGACCACAAATA-----	685
DB	110	GluSerProLysAlaArgGlyProArgProAsnValGluGlyArgAspGlyThrProGln	129
QY	686	-----ACTTGGCTGGTACTGTGACTGCTTT	712
DB	130	LysLysLysGlnCysAsnCysLysHisSerArgCysLeuLysLeuIleArgCysGlyPhe	149
QY	713	GCCAGTGGGACTTTTGCACAACTGCAATGTAATATTTGTCACAACTTGCAATCAT	772
DB	150	AlaSerGlyThrTyrcysAspGlyCysAsnCysValAsnCysPheAsnAsnValAspAsn	169
QY	773	GATATTCAAGCGTTAAAGCCATTAAGCATGCTTGGTAGAATCCAGAGCTTCCAG	832
DB	170	GluProAlaArgArgGlyAlaValGluAlaThrLeuGluArgAsnProPheAlaPheArg	189
QY	833	CCRAAAATTTGGGAGGCCAATTTGGCAAT-----	862
DB	190	ProLysIleAlaSerSerProHisGlyArgAspLysArgGluAspIleGlyGluVal	209
QY	863	---GTCAGCCCGCAGCACAAAGGTCGAATGCACTGCAGGAGGTCAGGCTCCCTGAAGAAT	919
DB	210	ValLeuLeuGlyLysHisAsnLysGlyCysHisCysLysLysSerGlyCysLeuLysLys	229
QY	920	TACTGCGAGTGCTATGAGGCCCAATATATGTTCTTCTTATTTGCAATGCATGTTGC	979
DB	230	TyrcysGlyCysPheGlnAlaAsnIleLeuCysSerGluAsnCysLysCysLeuAspCys	249
QY	980	AAAAATTATGAAGAAAGCCAGAACAGAACACACTAATGAGCATGCCAAAC-----	1030
DB	250	LysAsnPheGluGlySerGluGluArgGlnAlaLeuPheHisGlyGluHisSerAsnHis	269
QY	1031	-----TACATGCAG-----ACTGGAGTTTGGAGCGCAGCCAT	1063
DB	270	MetAlaTyrrLeuGlnGlnAlaAlaAsnAlaAlaIleThrGlyAlaValGlySerSerGly	289
QY	1064	TACCTGCCA-----CCAACCAATTTTCAGG-----	1090
DB	290	PheAlaProSerProAlaProLysArgGlyLysGlyGlnGlnIleLeuPheAsnGlnAla	309
QY	1091	CTTCCAAAGATTACGTACGATAGCGGCGCTTCTCTGATCATCTCCCTGGGAGGTGGAG	1150
DB	310	Ile-LysAspSerSerArgLeuSerHisPheProGlnValAsnGly-----	325
QY	1151	GCCACATGCGCCTGCTGCTGAGGAGAGAGCGGCGGAGAAACACACTGCTCCAAG	1210
DB	326	-----ArgThrGlyGlyProThrSerGlyThrSerProse	337
QY	1211	T 1211	
DB	337	r 337	
RESULT 9			
Q9CAV1			
ID	Q9CAV1	PRELIMINARY;	PRT; 601 AA.
AC	Q9CAV1;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Hypothetical 66.8 kDa protein.		
GN	T9J14.20.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLOMBIA;		
RX	MEDLINE=21016720; PubMed=11130713;		
RA	Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unselnd M.,		
RA	Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,		
RA	Deiseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,		
RA	De Simone V., Cholsne N., Artiguenave F., Robert C., Brottier P.,		
RA	Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,		
RA	Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,		
RA	Wiedelmann R., Drzonek H., Erfle H., Jordan N., Bangert S.,		
RA	Vezi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,		
RA	Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,		
RA	Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,		
RA	Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,		
RA	Cooke R., Laude M., Berger-Llauro C., Purnelle B., Masuy D.,		
RA	de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,		
RA	Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,		
RA	Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,		
RA	Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,		
RA	Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,		
RA	Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,		
RA	Pai G., Militscher J., Sellers P., Gill J.E., Feldhym T.V.,		
RA	Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,		
RA	Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,		
RA	Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,		
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,		
RA	Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,		
RA	Watanabe A., Yamada M., Yasuda M., Tabata S.,		
RT	"Sequence and analysis of chromosome 3 of the plant Arabidopsis		
RT	thaliana."		
RL	Nature 408:820-822 (2000).		
DR	EMBL; AC009465; AAG51411.1; --		
DR	InterPro: IPR005172; CXC.		
DR	Pfam: PF03638; CXC; 2.		
KW	Hypothetical protein.		
SQ	SEQUENCE 601 AA; 66786 MW; 10848E970D81E022 CRC64;		
Alignment Scores:			
Pred. No.:	4,75e-17	Length:	601
Score:	264.50	Matches:	81
Percent Similarity:	36.95%	Conservative:	28
Best Local Similarity:	27.46%	Mismatches:	101
Query Match:	6.92%	Indels:	85
DB:	10	Gaps:	12
US-09-743-237-3 (1-2134) x Q9CAV1 (1-601)			
QY	314	ATCTTTTGTCTCAGGAATCCTGTTTGCAGATGTCCTCCAGTCCAGGAAGTCCAGGATGCC	373
DB	255	AsnLeuSerCysSerSerLysValAlaAlaIleAspSerThrAlaGluAlaGluAspLys	274
QY	374	TCTGCTGTTCTCTTAAAGAAAGATCCACCAATGCTGATATGCCAATTAAGAGGGGC	433
DB	275	Glu-----AspLysAspLeuGlnPro-----SerGly	283
QY	434	ACACAAATGCTATGTATAGACAATCTAGAACAGAGAACTAAAGCACTCCATTTGGTT	493
DB	284	LysGlnArgSerValArgArgCysLeuThrPheAspMetGlySerHisLysArg	303

```

QY 494 CCTCAGTATCAAGATCAAAATAATATCTACAGTCAGAT-----GTC 535
Db 304 IleProLeuArgAspSerThrAsnAspLeuProLeuAspSerThrIleAsnLysAla 323
QY 536 CTTAAACCAATGACTGCTTTA-----GTAGGAGATTTTGGCCAGCA 577
Db 324 ProSerProGlnAsnCysLeuAspThrSerLysGlnAspThrAspGluIleLeuProIle 343
QY 578 TCAACAAATAATATCTACACACAACTTGAGGAGCCCTTA---CCATCGGTA--- 631
Db 344 ProArgThrIleGlyLeu-----HisLeuAsnGlyPheValAsnProSerValSer 360
QY 632 -----GTCAACGGGTCTCTCTTCCCTCGGGATCA----- 661
Db 361 SerGlyArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 380
QY 661 ----- 661.
Db 381 TyrAsnIleGluAspGluPheSerThrProValSerThrLysArgAspLeuValValPhe 400
QY 662 -----ACTCTTCAGGACCAACCAAAATAACTTTGGCTGGG----- 697
Db 401 SerAspValLysIleMetGluProProGluArgSerValGluGlyLeuPheCysAspGln 420
QY 698 -----TACTGTGACTGCTTTGCCAGTGGGGACTTTTGCACAAAC 736
Db 421 LeuMetAlaMetGluAsnArgTyrCysGluCysPheSerAlaGlyLeuPheCysGlyGlu 440
QY 737 ---TGCAATTGTATATTTGCAACAACCTTCATCATGATATTTGAACGGTTTAAAGCC 793
Db 441 ProCysSerCysGlnAsnCysPheAsnLysProIleHisGluAspLeuValMetLysSer 460
QY 794 ATTAAGGCATGCTTGTGTAACATCCAGAAGCTTTCAGGCCAAATTTGGAGGGCCAA 853
Db 461 ArgGluValIleLysAlaArgAsnProLeuAlaPheAlaProLysValSerThrSer 480
QY 854 -----TTGGCAATGTCAAG-----CCCCAGCACAAAC 880
Db 481 AspThrValIleAspLeuTrpValGluAsnSerLysThrProAlaSerAlaArgHisLys 500
QY 881 AAGGGTGCAACTGCAGGAGGTGAGCTGCTGAAGAAATTTACTGCGAGTCTATGAGGCC 940
Db 501 ArgGlyCysAsnCysArgLysSerGlyCysSerLysLysTyrCysGluCysPheMetMet 520
QY 941 CAATATGTTCTTCTATTTCAAATGCAATGCTGTTGCAAAAT 985
Db 521 GlyValGlyCysSerSerAsnCysArgCysMetGlyCysLysAsn 535

RESULT 10
Q9LW71 ID Q9LW71 PRELIMINARY; PRT; 356 AA.
AC Q9LW71
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Genomic DNA, chromosome 3, pl clone: MSL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Nakamura Y.;
RX MEDLINE=20277480; PubMed=10819329;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty pl and TAC

```

```

RT clones."
RL DNA Res. 7:131-135(2000).
DR ENBL: AB012247; BAB02682.1; -
DR InterPro: IPR005172; CXC.
DR InterPro: IPR000345; CytC_heme_bind.
DR Pfam: PF03638; CXC; 2.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
SQ SEQUENCE 356 AA; 39796 MW; 4FE0EA2D9DBFF58 CRC64;

Alignment Scores:      6.47e-17      Length:      356
Pred. No.:      262.50      Matches:      56
Score:      45.34%      Conservative:      17
Best Local Similarity:      34.78%      Mismatches:      47
Query Match:      6.86%      Indels:      41
DB:      10      Gaps:      5

US-09-743-237-3 (1-2134) x Q9LW71 (1-356)
QY 698 TACTGTGACTGCTTGGCCAGTGGGACTTTTGGCAACACTGCAATTGTAATAATTGTTGC 757
Db 79 TyrCysAspCysPheAlaSerGlyValValCysThrAspCysAspCysValAspCysHis 98
QY 758 AACAACTTGCATCATGATATTGAACGGTTTAAAGCCATTAAGGCATGCTTGTGTAGAAAT 817
Db 99 AsnAsnSerGluLysCysAspAlaArgGluAlaAlaMetValAsnValLeuGlyArgAsn 118
QY 818 CCAGAACCTTTCAGCCCAAAATTTGGAGGGCCAAATTTGGCAATGTCACAGCCCCAGCAC 877
Db 119 ProAsnAlaPheSerGluLys-----AlaLeuGlySerLeuThrAspAsnGln 134
QY 878 -----AACAAAGGTGCAACTGCAGAGG 901
Db 135 CysLysAlaAlaProAspThrLysProGlyLeuLeuSerArgGlyCysLysArg 154
QY 902 TCAGGCTGCCTGAAGAATTTACTGCGAGTGTATGAGGCCCAAAATTTATGTTCTTCTATT 961
Db 155 ThrArgCysLeuLysLysTyrCysGluCysPheGlnAlaAsnLeuLeuCysSerAspAsn 174
QY 962 TGCNAATGCATTTGTTGCAAAAATTTATGAAGAGCCCAAGCAAGCAAGCAACTAATGAGC 1021
Db 175 CysLysCysIleAsnCysLysAsnValSerGluAla----- 186
QY 1022 ATGCCAAACTACATGCAGACTGGAGGTTTGGAGGAGCCACTTACCTGCCACCAACGAAA 1081
Db 187 -----PheGlnProProAla--- 191
QY 1082 TTTTCA-----GGACTTCCAAGATTCAATGATGATGAGCGGCGCTTCTCATGATCATCTCC 1135
Db 192 PheSerAlaHisAsnSerProGlnValTyrArgArgArgArgArgGluLeuThrGlu 211
QY 1136 TGG 1138
Db 212 Trp 212

RESULT 11
Q92S22 ID Q92S22 PRELIMINARY; PRT; 896 AA.
AC Q92S22
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cysteine-rich polycomb-like protein.
GN CPPL.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
NCBI_TaxID=3847;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20345129; PubMed=10859345;
RA Cvitanich C., Pallisgaard N., Nielsen K.A., Hansen A.C., Larsen K.,

```

RA Pihakski-Maunbach K., Marcker K.A., Jensen E.O.;
 RT "CPPI, a novel type DNA-binding protein involved in the expression of
 a soybean leghemoglobin c3 gene."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:8163-8168(2000).
 DR EMBL: AJ010165; CAA09028.1; -
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;

Alignment Scores:
 Pred. No.: 2,06e-15 Length: 896
 Score: 248.50 Matches: 50
 Percent Similarity: 51.22% Conservative: 13
 Best Local Similarity: 40.65% Mismatches: 31
 Query Match: 6.50% Indels: 29
 DB: 10 Gaps: 5

US-09-743-237-3 (1-2134) x Q92S22 (1-896)

QY 698 TACTGTGCTGCTTCCAGTGGGACTTTTGCACAAC---TGCAATGTGTAATATTGT 754
 Db 487 TyrCysAspCysPheAlaAlaGlyThrTyrcysThrAspProcysAlaCysGlnGlyCys 506
 QY 755 TGCACAACACTTGCATGATATGAAGCTTTAAAGCCATTAAAGCATGTCTT---GGT 811
 Db 507 LeuAsnArgProGluTyr---ValGluThrValValGluThrLysGlnGlnLeuGluSer 525
 QY 812 AGAAATCCAGAACGCTTCCAGCCAAAAATTGGGAAGGCCAATTGGCCAATGCAAGCCC 871
 Db 526 ArgAsnProIleAlaPheAlaProLysIle-----ValGlnPro 538
 QY 872 -----CAG 874
 Db 539 ThrThrAspIleSerSerHisMetAspAspGluAsnLeuThrThrProSerSerAlaArg 558
 QY 875 CACAACAAGGGTGCACCTGCAGGAGCTCAGGCTGCTCGCTCAAGAAATTACTGCGAGTCTAT 934
 Db 559 HisLysArgGlyCysAsnCysLysArgSerMetCysLeuLysLysTyrCysGluCysTyr 578
 QY 935 GAGGCCCAAAATGTGTCTTCTATTGCAATGCAATGCTTGGCAAAATATGAAGAA 994
 Db 579 GlnAlaAsnValGlyCysSerSerGlyCysArgCysGluGlyCysLysAsnValHisGly 598
 QY 995 AGCCCAAGAA 1003
 Db 599 LysLysGlu 601

RESULT 12
 O23333
 ID O23333 PRELIMINARY; PRT; 658 AA.
 AC O23333;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 72.1 kDa protein.
 GN AT4G14770.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]

SEQUENCE FROM N.A.
 RP MEDLINE=9812113; PubMed=9461215;
 RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
 RA Bergkamp R., Dirke W., van Staveren M., Stiekema W., Drost L.,
 RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
 RA Wambutt R., Weitzenecker T., Pohl T.M., Terryn N.,
 RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
 RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kottler P.,
 RA Entlan K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
 RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
 RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,

RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
 RA Palme K., Benes V., Rechman S., Ansoerge W., Cooke R., Berger C.,
 RA Delsen M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
 RA Schueller C., Chludzisz N.,
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 Arabidopsis thaliana."
 RL Nature 391:485-488(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 297337; CAB10256.1; -
 DR EMBL: AL161539; CAB78519.1; -
 DR InterPro: IPR005172; CXC.
 DR Pfam: PF03638; CXC; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 658 AA; 72129 MW; 5BB99148E5E13A77 CRC64;

Alignment Scores:
 Pred. No.: 2.11e-15 Length: 658
 Score: 248.00 Matches: 92
 Percent Similarity: 35.07% Conservative: 49
 Best Local Similarity: 22.89% Mismatches: 143
 Query Match: 6.49% Indels: 118
 DB: 10 Gaps: 14

US-09-743-237-3 (1-2134) x O23333 (1-658)

QY 287 ATATCCGGAGAGCAACTTTGCAGATCTTCTGCTCAGAACTCTGTTCAGATGTC 346
 Db 219 AsnGlyValGluGlyGlnThrMetGlnHis----- 228
 QY 347 CCATGCTCCAGCACTAGAGGATGCTCTGCTGCTTCTTAAGAAAGATTCCACCCA 406
 Db 229 -----AspSerAsnLysGluProGluSerAlaAsnAlaIleProTyrGluValAsnSer 246
 QY 407 ATGGTGATATGCCAA-----TTGAAAGGGGGCACACAATGCTGTATAGAC 454
 Db 247 GlyValIleSerGlnAlaValSerLeuLeuHisArgGlyIleArgArgGlyCysLeuAsp 266
 QY 455 -----AATCTAGACAAGAACTA 475
 Db 267 PheGluMetProGlyAsnLysGlnThrSerSerGluAsnThrAlaAlaCysGluSer 286
 QY 476 AAAGCATCCATTGGTTCCTCAGTAT----- 502
 Db 287 SerSerArgCysValValProSerIleGlyLeuHisLeuAsnAlaIleLeuMetSerSer 306
 QY 503 CAAGATCAAAATATATCTACAGTCAGATGCTCCCTAAACCAATGACTGCTTTAGTAGG 562
 Db 307 LysAspCysLysThrAsnValThrGlnAspTyrSerCysSerAlaAsnIleGlnValGly 326
 QY 563 AGATTTTCCAGCATCAACAAAATAATCTCATTACACAACAACTTCAGGAGCCTTA 622
 Db 327 LeuGlnArgSerIleSerThr-----LeuGlnAspSerLeu 338
 QY 623 CCATCGGTAGTCAACGGGTCTGCTTCCCTCGGGATCAACTCTCCAGGACCAACA--- 679
 Db 339 AspGlnThrGluAsnGluIleArgGluAspAlaAspGlnAspValProValGluProAla 358
 QY 680 -----AAATAACTTTGGCT-----GGTACTGTGACTGCTTTGCCAGT 718
 Db 359 LeuGlnGluLeuAsnLeuSerSerProLysLysLysSerTyrCysGluCysPheAlaAla 378
 QY 719 GGGGACTTTGCG---AACAACCTGCAATGTAATAATTTGTGCAACAACCTTGCATCATGAT 775
 Db 379 GlyValTyrCysIleGluProCysCysIleAspCysPheAsnLysProIleHisGlu 398
 QY 776 ATTGAACGGTTTAAAGCCATTAAAGCATGCTCTTGTGTAGAAATCCAGAACTTCCAGCCA 835
 Db 399 AspValValLeuAlaThrArgLysGlnIleGluSerArgAsnProLeuAlaPheAlaPro 418
 QY 836 AAAATGGGAAGGCG-----CAATGGGCAAT----- 862


```

RL DNA Res. 7:131-135(2000).
DR EMBL: AB022223; BAB01253.1; -.
DR InterPro: IPR005172; CXC.
DR Pfam: PF03638; CXC; 2.
SQ SEQUENCE 695 AA; 76276 MW; 34BBA0E450F6BCE1 CRC64;

Alignment Scores:
Pred. No.: 6e-15 Length: 695
Score: 243.50 Matches: 131
Percent Similarity: 34.79% Conservative: 68
Best Local Similarity: 22.90% Mismatches: 173
Query Match: 6.37% Indels: 200
DB: 10 Gaps: 23

US-09-743-237-3 (1-2134) x Q9LUI3 (1-695)
QY 65 CCCGGGATCCAGAGCTCAGCGC-----GCTGGAGGAGCTGCG 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 ProAsnAspSerGluAlaPheArgCysLeuValAspLysIleSerSerGluArgArg 222
QY 104 GCTCTCGAGGCCCGCAGCGCCCGC-----CTGCAAGTGCCTTCTGCTCTC 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 PheCysAlaGlyValLysSerThrLysArgProAspIleAsnLysAspIleProAlaAsn 242
QY 155 GCTGCTACCCGCGCCAGCGCGCGGCTGTTTCCCGC----- 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 GlySerSerAsnGluAsnGlnProLeuAlaValLeuProThrAsnGluSerValPheAsn 262
QY 194 CTGGGGCGCTGGTCTCGCGAAG-----AGCTCCCAACCCGGCGCTCC 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 LeuHisArgGlyGly-MetArgArgCysLeuAspPheGluMetProGlyLysArgLys 282
QY 239 GCATGATCCAGTGAATCAAGTAAGTGAAGTGGTACTACTACAGTAATAATCCGNA 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 slyAsp-----IleValAspAspI 289
QY 299 GAAGCAACTTTCAGAATCTTCTGCTCAGGAATCCTGTTGCAAGTTCCTCATGGTCCAG 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 nGlnSerValCys-AspAsnAsnValAlaGlyLysSerSerSerCysValValProG 309
QY 359 GAAGTAGAGGATGCTC-----CTGCTGTTCTCTTAAGAAGATTCCAAACCATGGTG 412
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 lYileGlyLeuHisLeuAsnAlaValAlaMetSer-AlaLysAspSerAsnIleSerVal 328
QY 413 ATATGCCAATTGAAGGGGCGACAAATGCTATGTATAGACAATCTAGACAACAGANA 472
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 IleHisGlyTySerIleSerGlyGluIleGlnLysSerPheSerGlySerThrPro 348
QY 473 CTAAAGCACTCCATTTGTTCTCAGTATCAAGATCAA---AATAATTATCTACAGTCA 529
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 IleGlnSerGlnAspThrValGlnGluThrSerAspGlnAlaGluAsnGluProValGlu 368
QY 530 GATGCTCCCTAAACCAATGACTGCTTTAGPAGGAGATTTTCCAGCATCAACAATAA 589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 GluValProLys-----AlaLeuVal-----PheProGluLeu 379
QY 590 AATCTCATACACAACAACTTGAAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTC 649
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 AsnLeu----- 381
QY 650 CCCTCGGGATCAACTCTTCAGGACCACCAAAAATAACTTTGGCTGG----- 697
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 -----GlySerLeuLysLysLysMetArgLysSerGluGlnAlaGlyGluGlyGluSer 399
QY 698 -----TACTGTGACTGCTTTGCC 715
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 400 CysLysArgCysAsnCysLysLysSerLysCysLeuLysLeuTy-CysGluCysPheAla 419
QY 716 ACTGGGACTTTTGC-----AACAACTGCAATTTGTAATTTGTAACAACTGATCAT 772
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 AlaGlyValTyCysIleGluProCysSerCysIleAspPheAsnLysProIleHis 439
QY 773 GATATTGAACGGCTTTAAAGCCATTAAAGCATGCTTGGTAGAATCCAGAACCTTTCCAG 832

```

Search completed: April 21, 2003, 12:01:00
Job time : 156.5 secs

```

Db 440 GluGluThrValLeuAlaThrArgLysGlnIleGluSerArgAsnProLeuAlaPheAla 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 833 CCAAAATTT-----GGGAAGGGCCAATTTGGGC 859
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 ProLysValIleArgAsnAlaAspSerIleMetGluAlaSerAspAlaSerLysThr 479
QY 860 AATGTCAAGCCCGCAGCACAAAGGGTCAACTCAGGAGGTCAAGCTCGCTGAGAAAT 919
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 ProAlaSerAlaArgHisLysArgGlyCysAsnCysLysSerAsnCysMetLysLys 499
QY 920 TACTGCGAGTGTATCAGGCCCAAAATTTGTTCTTCTTATTTGCAAAATGCAATTTGCTTC 979
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500 TyrCysGluCysTyrGlnGlyGlyValGlyCysSerMetAsnCysArgCysGluGlyCys 519
QY 980 AAAAATTAT----- 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 ThrAsnValPheGlyArgLysAspGlySerLeuLeuValIleMetGluSerLysLeuGlu 539
QY 989 -----GAAGAAGCCCGAAGAACACACATAATGAGCATGCCAAACTACATGCGAGCT 1042
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 GluAsnGlnGluThrTyrGluLysArgIleAlaLysIleGlnHisAsnValGluValSer 559
QY 1043 GGAGGTTTGGAA----- 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 560 LysGluValGluGlnAsnProSerSerAspGlnProSerThrProLeuProProTyArg 579
QY 1055 -----GGCAGCCATTACCTGCCACCAACGAAATTT--- 1084
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 HisLeuValValHisGlnProPheLeuSerLysAsnArgLeuProProThrGlnPhePhe 599
QY 1085 -----TCAGGACTTCCAAGATTACGTCACGATAGCGGCTTCTCATGTCATCTCCTCG 1138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 LeuGlyThrGlySerSerPhe-----ArgLysProAsnSerAsp 613
QY 1139 GAGGTGGTGGAGCCACATCGCCTCGCTGCTCAGGAGAGAGAGCCGAGAAAGAA 1198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 614 -----LeuAlaGlnSerGlnAsnGluLysLys--- 622
QY 1199 CACTGCTCAAGTCCCTGGCAGAGCAGATGATCTCGGAGGAATTTGGAAGGTGCTTATCA 1258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 -----ProLeuGluThrValThrGluAspLysThrGluIleMetPro 636
QY 1259 CAGATTCTCCACACTCAGTTTAAATTAAGGGATTGAAATGGAGTAGAGTATAAGTCT 1318
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 GluIle----- 638
QY 1319 GAATGCATGTTGATTTTGTCTTAGTCTAGAAATCTCTAGTTTAGAAGGATGTTAGGGG 1378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 639 -----LeuLeuAsnSerProIleAlaAsnIle-----LysAlaIleSerPro 652
QY 1379 AACATGAGCTGGCTCTGCAGCAACAACAGCTCCCTCGCATCCCTGGCCGAGGAGAT 1438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 653 AsnSerLysArgValSerProGlnProGlySerSerGluSer-----GlySer 669
QY 1439 TTACTCAGAGCTCTCTGAAGATGTGCAAC 1468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 IleLeuArg-----ArgArgGlyAsn 676

```

